

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 06:15:54 ; Search time 3276.58 seconds
(without alignments)
2615.705 Million cell updates/sec

Title: US-09-319-156a-6
Perfect score: 635
Sequence: 1 ccctgatacttaacctctc.....tgaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estl:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	485.6	76.5	771	10	BI087886 602852690
2	450.8	71.0	689	12	AG121669 Pan trogl
3	437	68.8	494	9	AA781423
4	432.4	68.1	522	9	AW971553
5	423	66.6	653	12	AG033781 Pan trogl
6	415.8	62.5	490	9	AI598135
7	397.4	62.4	440	10	BE732673
8	396	62.4	543	10	BI963185
9	394.8	62.2	651	12	AG058970
10	393.2	61.9	701	12	AG126669
11	393	61.9	609	12	AG066901 Pan trogl
12	392.6	61.8	443	9	AA837267
13	387.2	60.9	712	12	AG0892947
14	386.8	60.9	436	9	AI128526
15	385.2	60.7	446	9	AI393478
16	384.6	60.6	641	12	AG036629
17	384.2	60.5	485	9	AW511366

C 18	375.2	59.1	556	9	AU158595
C 19	365.8	57.6	470	9	AI074704
C 20	364.8	57.4	679	12	AG076758
C 21	364.2	57.4	415	9	AI128496
C 22	360.6	56.8	777	12	AG030228
C 23	359.8	56.7	425	9	AI570707
C 24	359.8	56.7	431	9	AA552941
C 25	359.2	56.6	490	9	AA426511
C 26	359.2	56.6	619	12	AG133542
C 27	359.2	56.6	674	12	AG091649
C 28	355.4	56.0	701	12	AG096321
C 29	355	55.9	458	10	R76086
C 30	354.8	55.9	422	10	N53177
C 31	354.6	55.8	654	12	AG063042
C 32	353.8	55.7	665	12	AG088104
C 33	351.2	55.3	621	12	AG088991
C 34	351	55.3	433	9	AI379210
C 35	350.8	55.2	410	9	AA250958
C 36	350.6	55.2	1169	10	BM472049
C 37	347.6	54.7	424	10	R27412
C 38	341.2	53.7	446	9	AI288235
C 39	332.8	52.4	687	12	AG065253
C 40	331.8	52.3	692	12	AG040467
C 41	331.6	52.2	388	10	H01325
C 42	331.2	52.2	471	9	AA709471
C 43	330	52.0	569	12	AG080978
C 44	328.8	51.8	438	10	R77278
C 45	318.8	50.2	840	12	AG093377

ALIGNMENTS

RESULT 1
BI087886 771 bp mRNA linear EST 20-JUN-2001
LOCUS 602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:493894 5',
DEFINITION mRNA sequence.
ACCESSION BI087886
VERSION BI087886.1 GI:14506216
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 771)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgapb-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M11015 row: d column: 23
High quality sequence stop: 762.
location/Qualifiers
1. 771

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:493894"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 225 a 206 c 166 g 174 t

ORIGIN

Query Match 76.5%; Score 485.6; DB 10; Length 771;
 Best Local Similarity 89.9%; Pred. No. 7.7e-93;
 Matches 570; Conservative 0; Mismatches 49; Indels 15; Gaps 4;

```

Oy 1 cccgtatccttaacccctctgttaagttgtctctccagaatacaaaactgtaacta 60
    |||||||
Db 151 cccgtatccttaacccctctgttaagttgtctctccagaatacaaaactgtaacta 210
    |||||||
Oy 61 caaatgtcttcaaatgagacacagatgagatgacatgacatgacatgacatgac 120
    |||||||
Db 211 -----CAATGAGCCCAAGATGACGCTCAAGACATCAAGATCTACCGGAGACC 258
    |||||||
Oy 121 ctgagacgagctctgtagccatgctcagatgttaatgacatgagacacccctccag 180
    |||||||
Db 259 CTGAGACGGCGCTGATGACCGACGATCTGATGATGATGATGATGATGATGATG 318
    |||||||
Oy 181 gaatatcaactgcaaacacccctactatgcccccaatcagcgggaagcagttagaagcgt 240
    |||||||
Db 319 GAAATCTCAGCTGACACACTCTACTAGCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 378
    |||||||
Oy 241 catcagccaactcccccacacgcaacttggtttctcgtc-tgagaggggggagactagaaga 299
    |||||||
Db 379 CGTC-GGCAACCTCCCAAAAGCACTTGTGCTGTGATGAGTGGGAGCTAGAGA 437
    |||||||
Oy 300 cagagactagctgtagatctcctagcagcaagaatccctaaagcctcagcttggaaggtgac 359
    |||||||
Db 438 CAGGACTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497
    |||||||
Oy 360 tgcatacacccttaaacatgagggcttgcaacttagctcacacccgacccaatcagaagct 419
    |||||||
Db 498 CACATCCACCTTTAAACAGGGGCTTGCACACTGACACTGACCAATCAGAGAGACT 557
    |||||||
Oy 420 cactcaaaatgctcaatctagaacaaatagaaggta-aagaataagcaactcactctatgcc 478
    |||||||
Db 558 CACTTAAATGCTTAATTAAGCAAGACAGAGAGTACAAAGAAATAGCAATCATATTTGCC 617
    |||||||
Oy 479 tgaagacacagcggagagacaaagatcgagatatcaaacacccgacatcagcggagcagac 538
    |||||||
Db 618 TGAGAGCAGCAGAGAGAGGAGACATGATGGGATATTAAACCAAGCTTCGAGCGGAGAC 677
    |||||||
Oy 539 ggcacaccccttggtgctccctcctctgtatgagggcgtcgttttcaactatctaac 598
    |||||||
Db 678 GGCAACCCCTTTGGGCTCCCTCTTGTATGGAGCTGTATTCATGCTATTTCACT 737
    |||||||
Oy 599 ctataaatcttgcaactgtaaaaaaataaaaaa 632
    |||||||
Db 738 CTATTAAATCTTGCAACTGCAAAAACAAACA 771
    |||||||

```

RESULT 2
 AGI21669/c 689 bp DNA linear GSS 04-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
 DEFINITION AGI21669
 ACCESSION AGI21669.1 GI:16650834
 VERSION GSS: GSS (genome survey sequence).
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
 SOURCE BAC library clone:PTB-130M15.F.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
 1 (sites)
 Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of library PTB
 Unpublished
 JOURNAL 2 (bases 1 to 689)
 REFERENCE Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 AUTHORS
 TITLE Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimbeseqsc.riken.go.jp, URL:http://ngp-gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: -21M13

LIBRARY

Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.

FEATURES

Location/Qualifiers
 1..689

/organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-130M15.F"
 /sex="male"
 /cell_type="Lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"

BASE COUNT 164 a 159 c 176 g 189 t 1 others
 ORIGIN

Query Match 71.0%; Score 450.8; DB 12; Length 689;
 Best Local Similarity 88.3%; Pred. No. 1.8e-85;
 Matches 515; Conservative 0; Mismatches 58; Indels 10; Gaps 2;

```

Oy 1 cccgtatccttaacccctctgttaagttgtctctccagaatacaaaactgtaacta 60
    |||||||
Db 666 CCNTGATCTTTAAAGCTCTGTGATGATGATGATGATGATGATGATGATGATGATG 607
    |||||||
Oy 61 caaatgtcttcaaatgagacacagatgagatgacatgacatgacatgacatgac 120
    |||||||
Db 606 TTAATGCTTCTCAATGAGACACCTTATGACATGACATGACATGACATGACATGAC 547
    |||||||
Oy 121 ctgagacgagctctgtagccatgctcagatgttaatgacatgagacacccctccag 180
    |||||||
Db 546 CTGAGACGGCGCTGATGACCGACGATCTGATGATGATGATGATGATGATGATG 487
    |||||||
Oy 181 gaatatcaactgcaac-aacccctactatgcccccaatcagcgggaagcagttagaagcgt 239
    |||||||
Db 486 GAAATCTCAGCTGACACACTCTACTAGCTTCACTGACGACGAGCAAGCAGTTAGAGCGG 427
    |||||||
Oy 240 tcatcagccaactcccccacacgcaacttggtttctcgtcgttgaaggggggagactagaaga 299
    |||||||
Db 426 TCGTGGCCCAACTCCCAATGAGCACTTGCGGTTTCTGTTGAGGGGGGAGCTAGAGA 367
    |||||||
Oy 300 cagagactagctgtagatctcctagcagcaagaatccctaaagcctcagcttggaaggtgac 359
    |||||||
Db 366 CAGGACTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 307
    |||||||
Oy 360 tgcatacacccttaaacatgagggcttgcaacttagctcacacccgacccaatc----- 411
    |||||||
Db 306 CACATTCACCTTTAAACAGAGGCTTGCACACTTACCTCACACCCACCAATCAGCTAGTA 247
    |||||||
Oy 412 -agaagactcctaataatgtaataatgagcaaaatagaaggttaagaatgccaatc 470
    |||||||
Db 246 AAGAGGGCTCACTTAATAGCTTAATTAAGCAAAAACAGAGGTAAGAATAATGCCAATAT 187
    |||||||
Oy 471 ctatgcttgagagacagcggagagacaaagatcggagatatcaaacacccgacatcag 530
    |||||||
Db 186 TTATTGCTTGAGATGACGAGGAGAGACATGATGATGATGATGATGATGATGATG 127
    |||||||
Oy 531 ccggaacgcaaccccttggtgctccctcctctgtatggg 573
    |||||||
Db 126 CCACCAATGCTACCCCTTTGGGCTCCCTCCCTTTGTTATGG 84
    |||||||

```

RESULT 3
 AA781423/c

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NC1-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www.bio.llnl.gov/dbp/imagenet/imagenet.html
 Insert length: 1645 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 475.

Location/qualifiers
1. .494

a	111 c	133 g	141 t	1 others
---	-------	-------	-------	----------

d 111 c 133 g

ily 32.18; Freq. NO. 1.0E-02;
servative 0; Mismatches 36; Indels

GGCTTGCACCTTAGCTCACACCTGACCAATCAGAGAGCTCAC

Db 14 GCAACTGCAAAA 1

SOURCE	INITIAL
ORGANISM	Homo sapiens

1 (bases 1 to 5

M.E., Saeed, A.I., Shalov, V., Lee, N.H., Iacchini, J.C. and Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray
Unpublished (2000)

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel.: 301 838 3538

Tel: 301 838 3328
 Fax: 301 838 0208
 Email: johnq@tigr.org

Seq primer: Forward.

Location/qualifiers
1. .522

```

/NOE-VECTOR: pbaescrip1p3n11
125 a 115 c 134 g 148 t

```

Conservative	Mismatches	Indels	Gaps
0	41	2	2

CTGTGAGAGAGGGGACTGAGAGACAGGACTAGCTGATTTCTAGGCCGATTAGAAT 2833

Oy	335	ccctaagccttaagcttggaagatgactgtacatccaactcttaacaaatggggcttccaacttag	394
Db	282	cccttaaaccttaagcttgggaaggtgaccgcattccacctttttaaacaacggggcttgcacatttag	223
Oy	395	ctcacaccgcagcaaatcagaagatcctaactaaatgcttaattagtcanaaaataggaggttaa	454
Db	222	ctcacaccgcacccaattcagagagcttcaactaaatgcttaattgggcaaaaaaacagagaggttaa	163
Oy	455	agaaatagccaatcactatctgctcgtgagagcaacagcggaagagaaagaatcgagatatata	514
Db	162	agaatttagcccaattcatttatttgccttgacagacacagtggagggagcaagattgcacattata	103
Oy	515	aaccagagcatctgcagcgcaagcgcaaacccctttgggtccctccctttgtaatgtaggc	574
Db	102	aaccagagcatctgcagcgcaagcgcaaacccctttgggtccctccctttgtaatgtaggc	43
Oy	575	gctctgttttcaactatatttcaactatataaactctgcaact	616
	42	gctctgttttcaactatatttcaactatataaactctgcaact	1

RESULT	5				
LOCUS	AG033781/c				
DEFINITION	AG033781	653 bp	DNA	linear	GSS 01-NOV-2001
ACCESSION	AG033781	Pan troglodytes DNA, clone: PTB-008021.F,			genomic survey sequence.
VERSION	AG033781.1	GI:16560654			
KEYWORDS	GSS; GSS (genome survey sequence).				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1b:PTB Chimpanzee male BAC Library clone:PTB-008021.F.				

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
Pan troglodytes						
Eumayipota, Metacoza; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.						
1 (sites)						
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,						
Totoki, Y., Watanabe, H. and Sakaki, Y.						
BAC end sequences of library PTB						
Unpublished						
2 (bases 1 to 653)						
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,						
Totoki, Y., Watanabe, H. and Sakaki, Y.						

JOURNAL submitted (02-AUG-2001) Asso Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@sc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PMB This BAC endowment was generated during the Rad process and may have higher chance of clone tracking errors.

```

Sequencing: -21M13
LIBRARY
Vector      : pKS145
R.Site 1   : SacI
R.Site 2   : SacI
Location/Qualifiers
1. .653
FEATURES
source
```

```

/organism="Pan troglodytes"
/dd_xref="taxon:9598"
/clone="PTB-008021.F"
/sex="male"
/cell-type="lymphoblast"
/clone_lib="PTB chimpanzee"
BASE COUNT      145 a      155 c      178 g      173 t
ORIGIN
2 others

```

Query Match:	66.6%	Score 423;	DB 12;	Length 653;
Best Local Similarity:	88.0%	Pred. No. 1.3e-79;		
Matches 511; Conservative	0;	Mismatches 52;	Indels 18;	Gaps 4;

QY	2	cccttacctttaaaccctc-ttgttaaaatttcccttcccaaaacaaacgtttaaacta	60
Db	639	cccttttctttaaaccctcgtttgaagtgtctcttccaaatcaanaagctttaaagcta	580
QY	61	caaatgttcttccaatbtagagaccagaatbgaatcatalbactaaagatccacgtygacc	120
Db	579	CAAAATGGTTCTTCAAAATGGAGGCCCAAAATG-----CAGTCTCATGGACCC	535
QY	121	cttggaacgggctctgtatgcccattgtctccgaatgttaatacatltgaagcaacccctccgag	180
Db	534	CTGGACCGGGCTCTGTACCCATGCTCTGATGTATATGACAACGGAAGGC-GCCCTCTGAG	476
QY	181	gaaatccaaactgacaaaccctactatgcccgaattcaggggaagcaatftaagacgt	240
Db	475	AAAATCTCAACTGCACAAACCCCTACTACACCCAGTTTCAGCAGGAAGCAGTTAGAGTGT	416
QY	241	catcaagccaacacctccccaacagaacacttgggttttccgttctagaaggggggaactagaagac	300
Db	415	CGTTGGGCACACTCCCAACAGCACTTGGGTTTCCGTGTGAGAGGGGGGACTGAGAGAC	356
QY	301	aggactagcttgaatttctctagggccaagaaatccctbaagcctagctbgygaagtgaact	360
Db	355	AGGACTAGCTGAGATTTCCTAAAGCCAATTAAGAAATCCCTAAACCTACCTGGGAAGGTGACC	296
QY	361	gcatccaccctctaaacttgggctcttgaacttaactcaacaccgacccaatca-gagagct	419
Db	295	ACATCCACCTTTAAACATGGGGCTTGGAACTTACTCATCACCCAAACGTAAGAGAGACT	236
QY	420	caactaaactgttaatttaggcaaaatbaggagtgaaagaatataagccaatcatctatgtcct	479
Db	235	CACATAAATGCTATTATGGGCAAAAACAGAGAGTAAAGAAATAGCCATTCATCTTTGGCT	176
QY	480	gagagacagcggagaggaacaagaatcgggaatataaaccacagcatctcgagccggcgaacg	539
Db	175	GAGAGCACAGAGACAAGAGACAATGATCGAGATATATAAACCCAAAGCATTCGAGCCAGCAATG	116
QY	540	gcaaccccccttggctccctccctcttgatagaagcgctctg	580
Db	115	GCTACCTTCTTTGGTCCCTCCCTTCTGTATAGGAGCTCGG	75

RESULT	6
AI598135/c	
LOCUS	AI598135 490 bp mRNA linear EST 12-MAY-1999
DEFINITION	U11410.0.x1 NCI-CGAP-Brn25 Homo sapiens cDNA clone IMAGE:2167578 3'
	similar to contains PPR5.tl PPR5 repetitive element ; mRNA sequence.
ACCESSION	AI598135
VERSION	AI598135
KEYWORDS	AI598135.1 GI:4607183
SOURCE	EST.
	human.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 490)
NCI/NINDS-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CCGAP/ENGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.

CDNA library Preparation: M. Bento Soares, Ph.D., M. Fatma Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium at: www.bio.llnl.gov/bdnp/image/image.html
Insert length: 1407 Std Error: 0.00
Seq primer: -400P from Gibco

High quality sequence stop: 455
POLYA-No.

FEATURES	Location/Qualifiers
source	1. .490

```

/clone="IMAGE:2167578"
/clone.lib="NCI CGAP Brn25"
/issue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: p7T73d-Pec (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
stranded cDNA was primed with a Not I - oligo(dT) primer [5'
TGTATCCCAACATCAAGGCGAGCGCGCGATGCGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T73 vector.
Library is normalized, and was constructed by Bento
Library and M. Pothof-Baltes."

```

BASE COUNT	104 a	111 c	137 g	137 t	1 others
PRIGIN					

Query Match	65.5%	Score 415.8	DB 9	Length 490
Best Local Similarity	92.9%	Pred. No. 4.9e-78		
Matches 433; Conservative	0	Mismatches 33	Indels 0	Gaps 0

OY	126	ccggccgcgtatgcccatacgcctcctcgaatgataatgacatctgaagacccctcccgaggaat	180
Db	490	CCAGCCTGCTAGCCCTCTGTTCCGATGTTTAATGACATCAAAAGCACCCCTCTGAGGAAT	433
OY	186	ctcaactgacacaacccctactatgcccacattcagcgggaagcagttagaagcggatcalca	245
Db	430	CTCANTGCAACAACCCCTACCATGCCCATTTCCAGAGGAGCAGGTFAGACGGGCTCA	377
OY	246	gccaaactcccccaagagacttgggtttctctgtttgtagaggggagcagaaagacagagac	300
Db	370	GTCAAACCTCCCAACAGACACTGTGGTTTTCTGTTGAAGGGGGTACTGAAAGACAGGAC	311
OY	306	tactgtgatttcctctagcgcaacgagaatccctaaagcctagcttggaaagtactgacac	360
Db	310	TAGCTGGATTTCCTAGAGCGGCACTAAGAAATCCCTAAAGCTACTGAGGAAGTGACGGCTC	255
OY	366	caactctaaacaatggtgctctgcaactatgctccacacccgcgaacatcagaagatcactaa	420
Db	250	CACCTTTAAACAGGGGGCTTGCACCTTAGCTCAACCCGACCACAACTCAAGAGGCTCTCA	195
OY	426	aatgctaatatggcaaaatagaggttaagaataatagccaatcatcatcttgccttagagac	480
Db	190	AATGCTAATTAAAGCAAAAACAGGAGTTAAAGAAATAGCCAGTCACTTTCGCTGAGAGC	135
OY	486	acacgcgggaaggacaagatcgggatataaaacccagcatctcgagcggcaacggcaac	540
Db	130	ACAGCAAGAAGGAGACAATGATGGGATTAAAAACCCAGCATTCGAGCCAGCAACGGCTACG	715
OY	546	cccttctggctccctccctctgatatggcgctctggttttaactcatc	593
Db	70	CTCTTTGGGTCCTCCCTCTTTGTATAGGAGCCCTGCTTTTCACTCAATT	23

RESULT	7		
BE732673		440 bp	mRNA
LOCUS			linear
DEFINITION	BE732673 601571305F1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3925728 5'		EST 15-SEP-2000
ACCESSION	BE732673		
VERSION	BE732673.1		
KEYWORDS	GI:10146665		
SOURCE	EST.		
ORGANISM	human. Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 440)		

AUTHORS NTH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .440

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1IMAGE="3925728"
/clone_1lb="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10b (phage-resistant)"
/notes="Organ: Placenta; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1 kbp. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT	124 a	117 c	101 g	98 t
ORIGIN				

Query Match	62.68%	Score 397.4	DB 10	Length 440
Best Local Similarity	94.18%	Pred. No. 3.9e74		
Matches 413, Conservative	0	Mismatches 26	Indels 0	Gaps 0

OY	179	aggaatctcaacatgcacaacccctactatgcccacttcagcggagacagtttagagcg	238
Db	1	AGGAATTCCTGAGCTGSCANAACCTCTACTACGCCCCCAFTGACGAGAAAGACGTTAGAGCG	60
OY	239	gtcatcagcaacatccccaaacagcaacttgggtttccgtgtgagaggggagctagag	298
Db	61	GTCGCTGGCCCAACCTCCCCCAACACACTTACGTTTCTCTTTGAGAAAGGGGAGCTTGAG	120
OY	299	acagagcagcttgaatttcctcaggccaacggagaatccctaaagcctagctcggagaggtga	356
Db	121	ACAGGACTAGCTGGGATTTCTTAGGCTGACTAAGAAATCCTTAAGCCTTACCTGAGAGGTGA	180
OY	359	ctgcctccacactctaaacatgtygggcttgcactttagctcacaaccggacacaaatcagaaagc	418
Db	181	CCACATTCACCTTTAAACAGGGGGCTTGCAACTTAGCTGACACACTGACACATCAGAGAGC	240
OY	419	tcaactaaatgcttaattgagcaaaaatgaggtgaagaatagccaatcatctattggc	478
Db	241	TCACCTAAATTCCTAATTTGGCAAAAGACAGGGAGTAAACAAATATACCACATCATCTATTGGC	300
OY	479	tgaagagcacagcggagagagacaagatctggagataaaccagcattcgagccggcaac	538
Db	301	TGAGAGCACACAGGAGGAGGACATATGATGGGATTTAAACCCAAGCTTTCAGGCCGGCAC	360
OY	539	ggcaacccctttgggtccctccctctgtatagggcgctcgtttcacttattcaat	598
Db	361	GGCAACCCCTTTGGGTGGCTCCCTCCCTTGTATGGAGCCTGTGTTTCATGCTATTTCAC	420
OY	599	ctattaatatctgcaagc 617	
Db	421	CTATTAAATCTTGCAACTG 439	

RESULT	8			
BI963185/c				
LOCUS				
BI963185	543 bp	mRNA	linear	EST 23-OCT-2001

DEFINITION	1657c10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
ACCESSION	CDNA 3', mRNA sequence.
VERSION	BI663185
KEYWORDS	BI663185.1 GI:16337590
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 543)
	Melton,D., Brown,J., Kenly,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blisahn,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE	Endocrine Pancreas Consortium
JOURNAL	Unpublished (2000)
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@bldp.harvard.edu Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu) High quality sequence stop: 451.
FEATURES	Location/Qualifiers
SOURCE	1..543
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
	/sex="both"
	/tissue_type="Islets of Langerhans"
	/dev_stage="Adult"
	/lab_host="DH10B"
	/note="Organ: Pancreas; Vector: pSPORN1; Site.1: Not 1; Site.2: Sal 1; Starting library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot 0620. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT	116 a 123 c 147 g 156 t 1 others
ORIGIN	
Query Match	62.4%; Score 396; DB 10; Length 543;
Best Local Similarity	88.3%; Pred. No. 7.2e-74;
Matches 467; Conservative	0; Mismatches 51; Indels 11; Gaps 3;
0Y	74 aaatgagacacagatggagtcacatgactgaatccacgttgaccccttgacggcgcctg 133
Db	543 AAATGAGACCCCAAGATGACAGTCATGACTTAAGATCTACTGCGGACCTCGACGCGCTG 484
0Y	134 ctgagccatgtctcgagtttaatgacatggaagcaccctctccgaggaatctcaactg 193
Db	483 CTAGCCCA-GCTCCCACTTAATGATCAATCAAGCACCCTCCGAGGAAATCTCAACTG 425
0Y	194 cacaacccctactaagcccaaatcaaggagggaagcagttagaagcggtatcaagccaact 253
Db	424 CATGACACTTACTACGCCCCCAATTCAGCAGGAAGCAGTTAGAGTGCTGTCGGCCAACTT 365

OY	254	cccccaacgacttggtttcttccttcttagagggggagtgaagaacgactatcctgga	313
Db	364	CCCCAACAGCATTGTGGGTTTT-CTGTTTAGAGAGGGGACGTGAGACAGAGCATTA	306
OY	314	tctcttaaggccaagaataatccctaagacctgatgcggaaagtgaactcatcaccctca	373
Db	305	TTCCTTACGCCCACTAAGAATCCCTTAACCGCTTACGTGGGAAGTGACTACACCACCTTTA	246
OY	374	aacatlgggcgctgcacaattagctcacaccgccgaaccaat-----cagaagagctacta	424
Db	245	ACCACTAGAGCTTGCACAATTAGCTCACACCCGACANTAGAGGTAGTAAGAAGAGCTTGCTA	186
OY	425	aaatgtcataatgagcaaaaatagagggtcaagaagaatatcattcatctgcccggagag	484
Db	185	AAATCTATTATTAGGAAAAACAGAGGGGTAAAGAAAATATCCCACTCATCTATCCCTGACAG	126
OY	485	cacagcggagaggaagaagatcggagataaaaccagagcatcgcagccggcaagcgaaac	544
Db	125	CACAAAGGGGCGGGACAAATGATCAGAGATTAAACTAGAGCATTTCAAGCCACGACATGGCTAC	66
OY	545	cccccttggtccccctccctcttgtatgggcgcctcgtttccactcattt 593	
Db	65	COACTTGGGTGCCCTCCCATTTATTGGAGAGCTGTTTTCACCTGATT 17	
RESULT	9		
LOCUS	AG058970/c		
DEFINITION	Pan troglodytes DNA, clone: PTB-046A08.R, genomic survey sequence.		
ACCESSION	AG058970		
VERSION	AG058970.1 GI:16596431		
KEYWORDS	GSS: GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib=PTB Chimpanzee Male BAC library clone: PTB-046A08.R.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
REFERENCE	1 (sites)		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	BAC end sequences of library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 651)		
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shuhriro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:shimbes@gsc.riken.go.jp, URL:http://nbp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.		
PRIMERS			
SEQUENCING:	MJ3Rev		
LIBRARY	Vector : pKS145		
R.Site 1 :	Saci		
R.Site 2 :	Saci.		
Location/Qualifiers			
1..651			
/organism="Pan troglodytes"			
/db_xref="taxon:9598"			
/clone="PTB-046A08.R"			
/sex="male"			
/cell_type="lymphoblast"			
/clone_lib="PTB Chimpanzee Male BAC Library"			
BASE COUNT	141 a 160 c 159 g 191 t		
ORIGIN			

Query Match	Best Local Similarity	62.2%	Score 394.8;	DB 12;	Length 651;
Matches 503;	Conservative	0;	Mismatches 57;	Indels 14;	Gaps 6;
Oy	20	ttgttaagtgtgtctctccca-gaatcaaacgttaaactacaacatgtgtcttcaatg	78		
Db	613	TTGATAGGTTTGTCTCTCCCAAGANTCAAACGTATAAACAATAGTTCCTTCAATG	554		
Oy	79	gagcagcagatcgtatccatgac-taagatccaccgttgacccttgagaccgctctag	137		
Db	553	GAGCACGATACAGTGATGTCCTTAAGTTCCTCCGTGAACCCCTGGACGGCAGGTAG	494		
Oy	138	cccatgtctcgaatgttaatgacatitgaagaccctcccgaggaatcatctaactgaca	197		
Db	493	CCCATGCTCTATGTTAATGACATTGAAGGACCCCTCCGAGAAATCTCACTGACACA	434		
Oy	198	accctatactgcccacatcagcgggaagcagttagaacggtatcaagcaccctccc	257		
Db	433	ACCCCTGTACACCCCAATTCAGCAGGAACGATTTAGACAGTGTCTACGCCAACCTCCC	374		
Oy	258	aaccgacttgggttttccgtltbagagggggagctgagagacaggaatgacttgatttc	317		
Db	373	AACAGCACTTGGGTTTCTCTTGAGATGGGGAGCTAGACAGAACTACTGGATTTC	314		
Oy	318	ctagagcacaagaagaatccctaaagctagct-gggaaagtactgcatccactctaac	376		
Db	313	CTAGGCTGACTACAGAAATTCCTTAAGGCTGCTGGGGAAGATGATTGACCCACTTAAAC	254		
Oy	377	atggggcttgaacttagctcaaccccgacaaatag-----agagtcactaaa	427		
Db	253	ATGGGGCTTGTACTCACACTCACACCAAAATGATGTTACTAAAAAGGCTCACTTAAA	194		
Oy	428	tgctaatatagc-aaaaatagaggtlaaagaatatg-ccaatcatctatcttcctagagc	485		
Db	193	TACAATTTAGGCTAAACACGAGAGTAAAGAAATGTCAATATCATCTACGCTGAGACC	134		
Oy	486	acagcggagggagacaaggatcgggatatcaaacccaagcattcgagcggcagaacgcaac	545		
Db	133	ACAGTGGCAGGAGACAAATGTCGGGGTTATAAACCCAGCATTCGAGCAGGAGTGCACACC	74		
Oy	546	cccttgggtccccccttggatagggcgcttc	579		
Db	73	CGCTTGGGTCCTCCCTCCCATTTGATGAGGACTCT	40		
RESULT 10					
LOCUS AG126669/c					
DEFINITION	Ag126669	701 bp	DNA	linear	GSS 04-NOV-2001
ACCESSION	AG126669				
VERSION	AG126669.1	GI:1665834			
KEYWORDS	GSS: GSS (genome survey sequence).				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PRB Chimpanzee Male				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	1 (sites)				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
REFERENCE	Toxoki, Y., Metanabe, H. and Sakaki, Y.				
AUTHORS	BAC end sequences of library PRB				
TITLE	2 (bases 1 to 701)				
REFERENCE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
AUTHORS	Toxoki, Y., Metanabe, H. and Sakaki, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical				
COMMENT	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);				
	1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan				
	(E-mail: chimbesegsc.riken.go.jp, URL: http://ngp-gsc.riken.go.jp/				
	Tel:81-45-503-9111, Fax:81-45-503-9170)				
	Clones are derived from the chimpanzee BAC library PRB This BAC end				
	was generated during the Rtd process and may have higher chance of				

Clone tracking errors.										
PRIMERS										
Sequencing: -21M13										
LIBRARY										
Vector : pRS145										
R.Site 1 : SacI										
R.Site 2 : SacI.										
Location/Qualifiers										
1. .701										
/organism="Pan troglodytes"										
/db_xref="taxon:9598"										
/clone="PTB-137E19.F"										
/sex="male"										
/cell_type="Lymphoblast"										
/clone_11b="PTB Chimpanzee Male BAC Library"										
BASE COUNT	162 a	162 c	182 g	195 t						
ORIGIN										
Query Match	61.9%	Score 393.2;	DB 12;	Length 701;						
Best Local Similarity	88.1%;	Pred. No. 2.6e-73;								
Matches	465;	Conservative	0;	Mismatches	53;	Indels	10;	Gaps	3;	
QY	79	gagcaccagatggaatccatgaactaagatccacgcgttgagaccccttgacgcgcctctagc	138							
Db	651	GACCCCGGAGTCCATGCTCAAGACTGTAACCCCTGACGCTCTAC	592							
QY	139	ccatctccgatgttaatgacatgaaagcaccctcccgaagaatctcaatgcaca	197							
Db	591	CCATCTGCAATGTTAATGACATTGAAGCACCCTGCCGAGGAATTTCAACTGCACA	532							
QY	198	accctactatgcccccaattcagcggaaagcagtagagcggtcalcaagccaactcccc	257							
Db	531	ACCCCTACTATGCCCCGATTCAGCAGAGAGCAGTTAAAGTGGTCGTGGCCAAATCCCC	472							
QY	258	aacagcaacttggtttctcgttgagagggggagctgagagacaggactagctgatttc	317							
Db	471	AACACAGTGTGTTTCTGTTGAGAGGGGACTGAGAGACAGACACTGAGATTTC	412							
QY	318	ctagaccgaagaagaatcccttaagcttagcttggaagtgactgcatccactctaaaca	377							
Db	411	CTAGGCTGACTAAGANTCCCTTAAGACTAGCTGGAGAGGTGACACACTTCACCTTGAAACA	352							
QY	378	tggggcttgcaacttagctcacaccgcgaacaatc-----agagagctcaactaaatg	429							
Db	351	CGGGGTTGCAACTTAGCTCACATCCGACCAATGAGTTAAAGAAAGATAACTAAATG	292							
QY	430	ctaattaggcaaaaataggaggtlaaagaatatgccaatcatctatctgctgagagacag	489							
Db	291	CTAATTAGCAAAAACAGGAGGTAAAGAAATAGCCATATCATCTGAGATCAG	232							
QY	490	cggggggagcaagatctcggaatataaaaccgaagcatctagcgcggaaggaaccccc	549							
Db	231	CAGGAGGGACATGATCGGGGATATAAACCCAGGACATTCAGGACGACAAAGCTACCTCT	172							
QY	550	ttaggtccctccctcttgatgagcgctctgtttcaacttaattcac	597							
Db	171	TTTGGTCCCTCCCTTTGTATGGAGGCTGTGTTTCACCTATTAAAC	125							
RESULT	11									
LOCUS	AG066901	609 bp	DNA	linear	GSS 03-NOV-2000					
DEFINITION	Pan troglodytes DNA, clone: PTB-056L15.F, genomic survey sequence.									
ACCESSION	AG066901									
VERSION	AG066901.1 GI:16618703									
KEYWORDS	GSS; GSS (genome survey sequence).									
SOURCE	Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone: PTB-056L15.F.									
ORGANISM	Pan troglodytes									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.									
REFERENCE	1 (sites)									

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 609)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: schimpes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
COMMENT PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 609
/organism="Pan troglodytes"
/db_xref="taxon:9606"
/clone="PTB-056L15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 127 a 141 c 165 g 176 t
ORIGIN

Query Match 61.9%; Score 393; DB 12; Length 609;
Best Local Similarity 87.7%; Pred. No. 3e-73; Indels 12; Gaps 4;
Matches 478; Conservative 0; Mismatches 55;

QY 47 aaactgtaaactacaatctgtctcaaatgtgagcagacagatgagtcacatgataaga 106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 606 AAGCTGTAGACATACAAATCATCTTCAATGAGGCCCAATGAGTCATGATGATAAGA 547
QY 107 tcacacgtgagacccctgagccgctgctgtagccatgctcc-gaggttaataacattgaa 165
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 546 TCTACTGACAGACCTTGAGCAGCCCTGCTAGCTCAATGTTAATGGCTTGA 487
QY 166 ggcacc-ccctccgagagaaatcacaactgcaaacccctactatgcccattccagccgg 224
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 486 AGCCCTTCTCCCAAGAAATCTCAACTGCACACCCCTACTACACCCCAATTCACAGG 427
QY 225 aagcagttagagcgtcaltcagcacaactccccaacagcaactgggtttctcgtgtgaga 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 426 AAGCACTTAGACAGCAGTCACTGGCCAACTCCCAACAGCACTTGCTTTCATGTTGAGA 367
QY 285 ggggggagctagagagagagactgagttctcctagggcacaagaaatccctaagcct 344
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 366 TGGGGACTTTAGAGAGAGAGCTAGCTGGATTCTTA-GCTGACTAAGAAATCCCTAAGCCT 308
QY 345 agctggagaggtgactgcaactcctaacaatgagggcttgcaactatgctcaacccg 404
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 307 ACTGTGGAAGGTGACCACTTCACACTTTAAACACGGGGCTTTCACACTTACACACATG 248
QY 405 accaatca-----gagagctcaactaaatgctaaattaggaacaaatagaggttaa 455
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 247 ACCAATCAATAGTAGTAGAGCTCACTAAATGCTAATTAGGAAACAACAGAGAGGTAA 188
QY 456 gaataatgcaatcatctatctgctgagagacagcggggagacaaagtctgggtatata 515
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 187 GAAATAGCAATCACTTCTACTCTGAACACAGTGGAGGAGCAATGTTGGGATATA 128
QY 516 acccagagcctcagacggcgaacacccctctgaggtcccccctctgtagggagcg 575
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 127 ACCCAGGCACTTTGAGGCGACCAATGGCAACCCCTTTGGGTCCCTTCCTTTTATGGAG 68

QY 576 cctctg 580
|||||
DB 67 CTCGG 63

RESULT 12
AA837267 443 bp mRNA linear EST 31-MAR-1998
LOCUS od26b10.s1 NCI-CGAP GCBI Homo sapiens cDNA clone IMAGE:1369051
DEFINITION similar to contains PPR7.tl PPR7 repetitive element; mRNA
sequence.
ACCESSION AA837267 GI:2912466
VERSION AA837267.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bhrp/image/image.html
Insert Length: 1375 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
FEATURES
source
1. 443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1369051"
/clone_lib="NCI-CGAP_GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Gerald Marti (GBR). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGAGCGCGCTCAATTTTTTTTTTTT-3'
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 97 a 95 c 113 g 137 t 1 others
ORIGIN

Query Match 61.8%; Score 392.6; DB 9; Length 443;
Best Local Similarity 94.1%; Pred. No. 4e-73;
Matches 418; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 184 atctcaactgcaaacacctactatgcccacatcaggggaagcaagttagcggtcat 243
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 443 ATCTCAACTGCAACAACCCCTACTATAGCCCACTTACAGCGAAGCAAGTTAGTGATCAT 384
QY 244 cagccaacctccccaagagcacttgggtttctcctttagagagggggagcttagagacag 303
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
DB 383 CAGCCAACTTCAACCAACGACTTGGGTGTTTCTGTTGAGAGAGGGAGCTGAGACAGG 324

QY 304 actagctggaattcttagcgaacgaagaatccctaagccttagctgggaagtgaactgca 363
 Db 323 ACTAGTGAGATTTCCTAGTGGCGGATTAGAAATCCCTTAACCTAGTGGGAAGGTGACCGCA 264
 QY 364 tccaactctaacaatgggcttcgaacttagctcacaacccgaaccaatcagaagactact 423
 Db 263 TCCACTCTTAAACACGGGGCTTGCACCTTAGCTCACACCCACCAATCAAGAGACTACT 204
 QY 424 aaatgtctaattaggcaaaaataggaggtgaagaataagcaatcatctctgtctaga 483
 Db 203 AAATGTCTAATTAGGCAAAAACAGAGATGAAGAAATAGCCAAATCTATTGGCTTAGA 144
 QY 484 gcaacagggagggagcaacagatcgagatacaaacccagagcatcgaacggcaagca 543
 Db 143 GCACAGTGGGAGGACAGAGATTGCATATTAACCCAGCATTCGACCCAGC-ANAGCAA 85
 QY 544 ccccttgggtccctccctctgtatggcgctctgtttcaactatctactat 603
 Db 84 CCGCCTTGGGTCCCTTCCCTGTATGGAGCTCTGTTCACCTATTTCACCTATT 25
 Db 604 aaattgcaactgaaaaaanaa 627
 Db 24 AAATCTTGCACTGAAAAAANA 1

RESULT 13
 AOB92947/c 712 bp DNA linear GSS 10-NOV-1999
 LOCUS HS_3131_B2_E04_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3131 Col=8 Row=J, DNA sequence.

ACCESSION AOB92947
 VERSION AOB92947.1 GI:6349137
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3131 row: J column: 8
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 712.

FEATURES
 source
 1..712
 Location/Qualifiers

BASE COUNT 161 a 149 c 169 g 232 t 1 others
 ORIGIN
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_11b="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
 E-COLI DH10B"

Query Match 61.0%; Score 387.2; DB 12; Length 712;

Best Local Similarity 82.2%; Pred. No. 4,8e-72;
 Matches 484; Conservative 0; Mismatches 94; Indels 11; Gaps 3;

QY 56 aactacaattgtttctcaaatgagacacagatgagtgatcagatgaatcaccgtg 115
 Db 699 AAGTTACAAAGTTTCTNCNAAAGATCCCAAGATGCTAT-ACTCAAATCTCCGG 641
 QY 116 gacccctgagccgagcctgctagccagtcgcagtgtaagtacatgaaagccctc 175
 Db 640 ACCCTTGGGACCGGCTCGTATGTCATGCTTGCATGTGTGATATTAAGGACCCCTC 581
 QY 176 ccgaggaatctcaactcacaacccctactatgcccccaatlcagcgggaagcagtaga 235
 Db 580 CCGAGAAATCTCAAGTCATGACCTTAGT-TCACACAGTTCAGCAGGAAGCAGTTAGA 522
 QY 236 gcggtatcagcgaactcccccacagcaacttgggtttctctgttgaaggaggactga 295
 Db 521 GCGGCCCTTGGCCAACTCCCAATAGTACTTGGGTTTCTGTGAGAGGGGTTGCTGA 462
 QY 296 gagaacagactagctgattctcctagcgaacagagatccctaagctagctggagg 355
 Db 461 GAGACAGACTAGCTGATTTCTTAGCCGACACTAGATCCTTAGGCTTAGCTGGAGG 402
 QY 356 tgactgcatccactcctaacaatgggcttgcgaacttagctcacaccgacaaat----- 410
 Db 401 TGACTGCATCCACTTAAACACAGGGGCTTGCAACGTAGCTACACCCGACCAATGAGGT 342
 QY 411 -----cagagagctactaaatgctaattaggcaaaaataggaggtgaagaataagccaa 466
 Db 341 AGTAAAGAGAGCTCCTAAATGCTAATTAGGCAAAAACAGGAATGAAGAAATAGGCCAA 282
 QY 467 tcaatctatgctctgagagcaacagcggagagagagatcgagatacaaacccagagcat 526
 Db 281 TCATCTATCACTGAGACACAGGGGAGGAGACATGATGATATTAACCCAGGGGCTT 222
 QY 527 cgagcggcaacggcaaccccttgggtccctccctctgtatggcgctctgttca 586
 Db 221 CTACCGGCAACGGCTACCTCTTGGGTACCCCTTGTATGGAGAGCTCTGTTCAT 162
 QY 587 ctctattgctctctatcaatctgcaactgcaacgaaaaaanaaaaaa 635
 Db 161 CTCTATTAACTTGCNAAAGACACAAAACCAACCAACCAACCAANA 113

RESULT 14
 A1128526/c 436 bp mRNA linear EST 27-OCT-1998
 LOCUS g651h10.x1 Soares_placenta_8c9weeks_2bHP8to9w Homo sapiens cDNA
 DEFINITION clone IMAGE:1714147 3' similar to contans PIR7.b1 PIR5 repetitive
 element ;, mRNA sequence.

ACCESSION A1128526
 VERSION A1128526.1 GI:3597040
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 436)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert length: 720 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 428.

FEATURES
 source
 1..436
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"

QY 301 aggaactagctgattctccctagagcccaagaaatccctaaagcttagctgggaaggctgact 360
|||||
Db 2611 aggaactagctgattctccctagagctgactaaagaaatccctaaagcttagctgggaaggctgacc 2670
QY 361 gcatccacaccttaaacatgagggtctgcaacttagctcacacccgcaacaaatcagaagctc 420
|||||
Db 2671 acatccacctttaaacaagggtctgcaacttagctcacacccgcaacaaatcagaagctc 2730
QY 421 actaaacgctaatagagcaaaaatagaggtaaagaataagccatcatctatctgctg 480
|||||
Db 2731 actaaacgctaatagagcaaaaatagaggtaaagaataagccatcatctatctgctg 2790
QY 481 agagcacagcgaggaggaacaaagatcgggatatataaacccagagcaltcgagccggcaagc 540
|||||
Db 2791 agagcacagcgaggaggaacaaatgacgagatatataaacccagagcaltcgagccggcaagc 2850
QY 541 caacccctctgggtgccctccctctgtatgagggtctgttttaccatattacact 600
|||||
Db 2851 caacccctctgggtgccctccctctgtatgagggtctgttttaccatattacact 2910
QY 601 attaatcttgcaactggaataaaaaa 635
|||||
Db 2911 attaatcttgcaactggaataaaaaa 2945

RESULT 2
US-08-686-878A-50
: Sequence 50, Application US/08686878A
: Patent No. 5708157
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavalley, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Evans, Cheryl
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/686,878A
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 279 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-686-878A-50

Query Match 36.8%; Score 233.6; DB 1; Length 279;
Best Local Similarity 88.1%; Pred. No. 7,6e-66;
Matches 245; Conservative 1; Mismatches 32; Indels 0; Gaps 0;
QY 338 actgcatccaccttaaacatgagggtctgcaacttagctcacacccgcaacaaatcagaag 417
1 |||||
Db 1 RCGACATCCACCTTTAAACACGGGGTGTGCAANNAAGTACACTTGACCAATCGAGAG 60
QY 418 ctcaactaaatgctaatagcaaaaatagagggttaagaataagccatcatctatctg 477
|||||
Db 61 NTCANTAAATGATTAATTTNGCCAAAACAGAGAGTAAAGAAATAGCCAATCATCTATTGC 120
QY 478 ctgaagcacagcgaggaggaacaaagatcgggatatataaacccagagcaltcgagccggcaa 537
|||||
Db 121 CTGAGAGCAGACAGAGAGGAGCAATGATCGGATTTAAACCAAGTTTNGAGCCGCCAA 180
QY 538 cggcaacccctctgggtgccctccctctgtatgagggtctgttttaccatattac 597
|||||
Db 181 CGGCAACCCCTTGGGTGCCCTCCCTCTGTATGGAGCTGTGTTTCATGCTATTTCAN 240
QY 598 tctatgaactctgcaactggaataaaaaa 635
|||||
Db 241 TTTATTAAATTTTGCACTGCCAAAAA 278

RESULT 3
US-08-721-489-4
: Sequence 4, Application US/08721489
: Patent No. 5786465
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavalley, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/721,489
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 279 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-721-489-4

Query Match 36.8%; Score 233.6; DB 1; Length 279;
Best Local Similarity 88.1%; Pred. No. 7.6e-66;
Matches 245; Conservative 1; Mismatches 32; Indels 0; Gaps 0;

QY 358 actgcatccaccccttaaacatgaggtgctgcaactagctcaccaccacatcaagag 417
DB 1 RCCGATCCACCTTTAAACAGGGGNTTGCANANAAGATNACCTTGACCATCAGAG 60
QY 418 ctccataatgctaatgaagcaaaatagaggtlaagaatagccaatcatctatgc 477
DB 61 NTCNTAAATGATNATNTNGCAAAACAGAGSTAAAGAAATGCCATCATCTATTGC 120
QY 478 ctgagagcacagcgagaggaacaaagatcggaatataaacaccagcatctcgagccgag 537
DB 121 CTGAGAGCACAGCAGAGGACATGATCGGATATTAACCCAGTTTNGAGCCGAG 180
QY 538 cggagaccccttgggttccctcccttctgtatgagcgctctgttccactctaac 597
DB 181 CGGCAACCCCTTGGGTCCCTCCCTTGTATGGGAGCTTGTTCATGCTATTTCAN 240
DB 598 tctataatctgcaactgaaataaaaaa 635
DB 241 TTTATTAATTTTCCAACTGCAAAAAA 278

RESULT 4

US-08-691-563C-46
Sequence 46, Application US/08691563C
Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: Herve PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BEDIN
APPLICANT: Glaucia PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38568
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-08-691-563C-46

Query Match 20.2%; Score 128; DB 3; Length 1859;
Best Local Similarity 62.7%; Pred. No. 2.2e-31;
Matches 237; Conservative 0; Mismatches 130; Indels 11; Gaps 2;

QY 1 cccgtatccttaacctctgttaagtgtgtctctccagaatcaaaactgttaacta 60
DB 1446 CCCGTATTTTAAACCTCTTGTCAAAATGTTCTTCTAGAGTCAGGCCATCACTA 1505
QY 61 caatgtcttcaaatgagagcaccagatgagtcctagtaagatccaccgtagcc 120
DB 1306 CAGATGGTCTTACAAATGGAACCCCAATGAGCTCAACTATCACTTCTATGAGACCC 1565
QY 121 ctgagccgacctgtacccatgctccgagttaatgacatggaagcaccctccag 180
DB 1566 CTATACCAACGCCCTGGGCC-----TTTCACTGGCCCTAAAGAGTTCCTCTGAG 1616
QY 181 gaaatcacaactgacaacccctacatgcccacatcagcggagagcttagagcgt 240
DB 1617 GACACTACACATGACAGGCCCCCATCTTGGCCCTATCCAGAAAGACATGACAGCAGT 1676
QY 241 catcagcacaactcccaacagcaactgggttctccgtgtgagagggagcttagagac 300
DB 1677 CATTGCCCA--TTCCCAAGACAGCTGGGGTGTCCGTTAGAGTGGGATTAAGAGT 1734
QY 301 agagctagctgattctcctagggcacaagaaatccctaaagctagctggaagtgact 360
DB 1735 GAAGCCAGCTGTGACTTCTGGGTGGGGTGGGACTTTGGAGAACTTTGTGTACTAAG 1794
QY 361 gcaatccacttaaacat 378
DB 1795 GATTGTAATGCAACAAT 1812

RESULT 5

US-09-078-294-4
Sequence 4, Application US/09078294
Patent No. 6265211

GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies C01
CURRENT APPLICATION NUMBER: US/09/078, 294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 80246
TYPE: DNA
ORGANISM: Nucleotide sequence of NC-contlg
US-09-078-294-4

Query Match 19.8%; Score 125.8; DB 4; Length 80246;
Best Local Similarity 67.2%; Pred. No. 9.3e-30;
Matches 213; Conservative 0; Mismatches 92; Indels 12; Gaps 2;

QY 1 cccgtatccttaacctctgttaagtgtgtctctccagaatcaaaactgttaacta 60
DB 58888 cccgtatctttaaactctctgttcaaatgttctccttagagcagagcatcaacta 58947
QY 61 caatgtcttcaaatgagagcaccagatgagtcacatgactaagatccaccgtgagcc 120
DB 58948 cagatgacttaacaatgtaaccccaatgagctcaactaacaactctgtgagagcc 59007
QY 121 ctgagccgacctgtacccatgctccgagttaatgacatggaagcaccctccag 180
DB 59008 ctgagcagaccgctgagc-----ttcaatgcccctaaagagagctccctctgag 59058
QY 181 gaaatcacaactgacaacccctacatgcccacatcagcggagagcagcttagagcgt 240
DB 59059 gacactacaactgagggccctcttccaccctatccagcaggaagtagtacagcgt 59118

FILED DATE: 05 JUN 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/011,398
FILED DATE: 29 JAN 1993
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/160002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2417
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
--08-464-051-1

Query Match 4.9%; Score 31.4; DB 1; Length 2417;
Best Local Similarity 54.9%; Pred. No. 3.4;
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 521 ggcattcgagccgcaacgcaaccccttggtccctctgtatgagcgctctg 580
DB 2305 ggcctctgaccccttacccttggaagcccttgagacacttaccctggcctgttgacatt 2364
QY 581 tttaactctatctcaactctatctcaactctgcaactgaaaaa 633
DB 2365 tctatattttaaataatttaactacccttaattactttaaaaaa 2417

Search completed: June 20, 2002, 09:35:17
Job time: 8227 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 07:21:10 ; Search time 443.18 Seconds
(without alignments)
2460.040 Million cell updates/sec

Title: US-09-319-156a-6
Perfect score: 635
Sequence: 1 cccgtatcttaactcctc.....tgaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_032802.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	100.0	635	19 AAV43215	Multiple sclerosis
2	635	100.0	635	20 AAX29702	Clone CL6-3' from
3	617.4	97.2	2030	21 AAV63826	Nucleotide sequenc
4	545	85.8	1329	19 AAV43219	Multiple sclerosis
5	545	85.8	1329	20 AAX29704	Clone 5M6 from MSR
6	533.8	84.1	2946	20 AAX77526	Human secreted pro
7	533.8	84.1	2946	21 AAX77526	Human secreted pro
8	532.2	83.8	2781	22 AAX55650	Nucleotide sequenc
9	525.8	82.8	1136	20 AAX25660	Human endogenous r

10	525.8	82.8	1136	21 AAX59210	3' pol gene and 3'
11	522.6	82.3	2782	20 AAX25661	Human endogenous r
12	522.6	82.3	2782	21 AAX59211	5' non coding, 3'
13	522.6	82.3	2782	22 AAX20069	HERV-W envelope pr
14	511.4	80.5	7582	20 AAX25665	Complete human end
15	511.4	80.5	7582	21 AAX59215	Human endogenous r
16	499.6	78.7	1894	22 ABA45822	Human breast cell
17	499.6	78.7	1894	22 ABA56337	Human foetal liver
18	499.6	78.7	1894	22 ABA25978	Probe #4444 for ge
19	499.6	78.7	1894	22 AAK04516	Human brain expres
20	499.6	78.7	1894	22 AAK30018	Human bone marrow
21	499.6	78.7	1894	22 AAL14608	Probe #4541 for ge
22	499.6	78.7	1894	22 AAL35980	Probe #4666 used t
23	499.6	78.7	1894	22 AAL04422	Probe #4413 used t
24	495.6	78.0	3372	20 AAX25663	Human endogenous r
25	495.6	78.0	3372	21 AAX59213	Partial pol gene a
26	493.2	77.7	2052	22 ABA08802	Human receptor Tyr
27	481.4	75.8	849	22 AAS31000	Human diagnostic a
28	479.2	75.5	3831	23 AAS71727	DNA encoding novel
29	479.2	75.5	5154	23 AAS67609	DNA encoding novel
30	479.2	75.5	8279	23 AAS76474	DNA encoding novel
31	479.2	75.5	8294	23 AAS84209	DNA encoding novel
32	469.4	73.9	2942	23 AAS77313	DNA encoding novel
33	467	73.5	7466	23 AAS68626	DNA encoding novel
34	461.6	72.7	2527	22 AAH17481	Human cDNA sequenc
35	460.8	72.6	583	23 AAS83931	DNA encoding novel
36	453.4	71.4	1115	23 AAS76196	DNA encoding novel
37	451	71.0	3903	22 AAH46211	Phosphoribosyl pyr
38	449	70.7	1364	22 AAS63154	Human purified sec
39	448.2	70.6	2300	23 AAS92680	DNA encoding novel
40	446.2	70.3	1393	23 AAS31002	Human diagnostic a
41	437.8	68.9	893	22 AAS65964	DNA encoding novel
42	432.8	68.2	1666	23 AAS92500	DNA encoding novel
43	422	66.5	1165	23 AAS72721	DNA encoding novel
44	388.4	61.2	2385	22 AAI99411	Human excretory re
45	388.4	61.2	2385	22 AAI63761	Human kidney relat

ALIGNMENTS

RESULT	1	
ID	AAV43215	standard; cDNA; 635 BP.
XX		
AC	AAV43215;	
XX		
DT	29-DEC-1998	(first entry)
XX		
DE	Multiple sclerosis associated retrovirus fragment 4.	
XX		
KW	Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;	
KW	999 gene; env gene; rheumatoid arthritis-associated virus; ss.	
XX		
OS	Multiple sclerosis associated retrovirus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..234
FT		/*tag- a
FT		/product= "Encodes protein AAV71067"
XX		
PN	WO9823755-A1.	
XX		
PD	04-JUN-1998.	
XX		
PF	26-NOV-1997;	97WO-IB01482.
XX		
PR	26-NOV-1996;	96US-0756429.
XX		
PA	(INMR) BIO MERIEUX.	
XX		
XX	Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;	
PI	Mandrand B, Paranhos-Baccala G, Perron H;	

XX WPI: 1998-322732/28.
DR P-PSDB: AAW71067.
XX New nucleic acid from retroviruses - useful for diagnosis,
PT prevention and treatment of, e.g. multiple sclerosis
XX
PS Disclosure: Page 183; 286pp; English.
XX
CC The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) genomic fragment used in the method of the
CC invention. The invention provides complete or partial genomic
CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC polypeptides encoded by these genes. The invention also provides
CC antibodies raised against the polypeptides. The genomic sequences,
CC polypeptides and antibodies are also claimed useful for diagnosing
CC infection by MS and rheumatoid arthritis-associated viruses, and also
CC for prevention and treatment of infection with these viruses.
XX

Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 other;

Query Match 100.0%; Score 635; DB 19; Length 635;
Best Local Similarity 100.0%; Pred. No. 2.2e-187;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgtatctttaaaccctctgttaagttgtctctccccaagaatcaaaactgtaaacta 60
DB 1 cccgtatctttaaaccctctgttaagttgtctctccccaagaatcaaaactgtaaacta 60
QY 61 caaatgtcttcaaatgagacacacagatgagtcacatgaactaacacgctgagacc 120
DB 61 caaatgtcttcaaatgagacacacagatgagtcacatgaactaacacgctgagacc 120
QY 121 ctgagacgagcctgtagccacatgctcgaatgttaatgaatgaagcaccctccgag 180
DB 121 ctgagacgagcctgtagccacatgctcgaatgttaatgaatgaagcaccctccgag 180
QY 181 gaaatctcaactgacaaacccctactatgcccacatcagcgaggagacagtcgagc 240
DB 181 gaaatctcaactgacaaacccctactatgcccacatcagcgaggagacagtcgagc 240
QY 241 catcagccaactccccaacagcactgtgttctctgttgaaggggagctgagaaac 300
DB 241 catcagccaactccccaacagcactgtgttctctgttgaaggggagctgagaaac 300
QY 301 aggaactagctgatttccctaggaacgaagaatccccaagcctgagcgagagctgact 360
DB 301 aggaactagctgatttccctaggaacgaagaatccccaagcctgagcgagagctgact 360
QY 361 gcatccacaccttaaacatgagggtctgcaactagctcacacccgacacatcagaagagtc 420
DB 361 gcatccacaccttaaacatgagggtctgcaactagctcacacccgacacatcagaagagtc 420
QY 421 actaaatgcttaattagcaaaaatagaggtlaaagaatagccaatcattatgctcgt 480
DB 421 actaaatgcttaattagcaaaaatagaggtlaaagaatagccaatcattatgctcgt 480
QY 481 agagacagcgaggaggaagacagatcgagataaaacccaagcattcgagccgagaaagg 540
DB 481 agagacagcgaggaggaagacagatcgagataaaacccaagcattcgagccgagaaagg 540
QY 541 caaaccccttgggtccctccctctgtatagggcgctctgtttcactatcttcaactct 600
DB 541 caaaccccttgggtccctccctctgtatagggcgctctgtttcactatcttcaactct 600
QY 601 attaaatcttgcaactgaaaaaataaaaaa 635
DB 601 attaaatcttgcaactgaaaaaataaaaaa 635

ID AAX29702 standard; DNA; 635 BP.
XX
AC AAX29702;
XX
DT 08-JUN-1999 (first entry)
XX
DE Clone CL6-3' from MSRV-1.
XX
KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KW rheumatoid polyarthritis; ss.
XX
OS Multiple sclerosis related virus type 1.
XX
PN FR2765588-A1.
PD 08-JAN-1999.
XX
PF 07-JUL-1997; 97FR-0008816.
XX
PR 07-JUL-1997; 97FR-0008816.
XX
PA (INMR) BIO MERIEUX.
XX
DR WPI: 1999-098275/09.
DR P-PSDB: AAW9552.
XX
PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis
PS Claim 1: Page 36-37; 83pp; French.
XX
CC This sequence represents clone CL6-3' from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with
CC rheumatoid polyarthritis.
XX
SQ Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 other;

QY 1 cccgtatctttaaaccctctgttaagttgtctctccccaagaatcaaaactgtaaacta 60
DB 1 cccgtatctttaaaccctctgttaagttgtctctccccaagaatcaaaactgtaaacta 60
QY 61 caaatgtcttcaaatgagacacacagatgagtcacatgaactaacacgctgagacc 120
DB 61 caaatgtcttcaaatgagacacacagatgagtcacatgaactaacacgctgagacc 120
QY 121 ctgagacgagcctgtagccacatgctcgaatgttaatgaatgaagcaccctccgag 180
DB 121 ctgagacgagcctgtagccacatgctcgaatgttaatgaatgaagcaccctccgag 180
QY 181 gaaatctcaactgacaaacccctactatgcccacatcagcgaggagacagtcgagc 240
DB 181 gaaatctcaactgacaaacccctactatgcccacatcagcgaggagacagtcgagc 240
QY 241 catcagccaactccccaacagcactgtgttctctgttgaaggggagctgagaaac 300
DB 241 catcagccaactccccaacagcactgtgttctctgttgaaggggagctgagaaac 300
QY 301 aggaactagctgatttccctaggaacgaagaatccccaagcctgagcgagagctgact 360
DB 301 aggaactagctgatttccctaggaacgaagaatccccaagcctgagcgagagctgact 360
QY 361 gcatccacaccttaaacatgagggtctgcaactagctcacacccgacacatcagaagagtc 420
DB 361 gcatccacaccttaaacatgagggtctgcaactagctcacacccgacacatcagaagagtc 420
QY 421 actaaatgcttaattagcaaaaatagaggtlaaagaatagccaatcattatgctcgt 480

```

Db 421 actaaatgctaattagcgaataaagagagtaagaatagccaatcatctatgctg 480
      |||
Qy 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
      |||
Db 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
      |||
Qy 541 caaccccttgggtccccccttctgtatcggtctgttttctactatctcactc 600
      |||
Db 541 caaccccttgggtccccccttctgtatcggtctgttttctactatctcactc 600
      |||
Qy 601 attaatcttgcactgcaactgcaaaaaaa 635
      |||
Db 601 attaatcttgcactgcaactgcaaaaaaa 635
      |||

RESULT 3
AAAG3826
ID AAA63826 standard; DNA: 2030 BP.
XX AAA63826;
XX
DT 04-DEC-2000 (first entry)
XX
DE Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KM MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;
KM ss.
XX
OS Multiple Sclerosis retrovirus 1.
XX
FH Key Location/Qualifiers
FT CDS 1..1629
FT sig_peptide /tag= a
FT /note= "Contains one termination codon"
FT /tag= b
FT CAAT_signal 1800..1807
FT /tag= c
FT CAAT_signal 1858..1864
FT /tag= d
FT TATA_signal 1906..1911
FT /tag= e
FT polyA_signal 1996..2002
FT /tag= f
XX
PN MO200047745-A1.
XX
XX 17-AUG-2000.
XX
XX 15-FEB-2000; 2000MO-IB00159.
XX
XX 15-FEB-1999; 99EP-0420041.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX
XX WPI; 2000-506097/45.
XX
XX P-PSDB; AAB08195.
XX
XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
XX retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
XX biological sample
XX
XX Disclosure; Fig 2; 23pp; English.
XX
XX The present sequence represents the nucleotide sequence corresponding
XX to the 3' env region and long terminal repeat sequences from clone
XX C16 of Multiple Sclerosis retrovirus (MSRV-1). The specification
XX describes a long terminal repeat (LTR)-RUS region which encodes the
XX expression of a MSRV-1 protein. This is unusual for LTRs, in
XX particular in the RUS region. The sequence includes CAAT and TATA

```

```

CC signals which are present in the U3 and R regions and are not directed
CC towards the CDS indicated in the features table. Probes and antibodies
CC to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
CC are used to detect the presence of MSRV-1 retrovirus in a biological
CC sample.
XX
SQ Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 other;

Query Match          97.2%; Score 617.4; DB 21; Length 2030;
Best Local Similarity 98.3%; Pred. No. 1..2e-181;
Matches 624; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 cccgtatctttaaacccttggtaagtgtgtcttccagaaatcaaaactgtaaaacta 60
Db 1396 cccgtatctttaaacccttggtaagtgtgtcttccagaaatcaaaactgtaaaacta 1455
Qy 61 caaatgtcttcaaatgagacaccagatgagtcataagataccacgttgagacc 120
Db 1456 cagatgtcttcaaatgagacaccagatgagtcataagataccacgttgagacc 1515
Qy 121 ctggaccggtctgtatgccatgtccgatgttaattgaattgaagcaccctccgag 180
Db 1516 ctggaccggtctgtatgccatgtccgatgttaattgaattgaagcaccctccgag 1575
Qy 181 gaaatcctaactgacacacccttactatgcccccaattcagcggaagcagttagcggt 240
Db 1576 gaaatcctaactgacacacccttactatgcccccaattcagcggaagcagttagcggt 1635
Qy 241 catcagccaactccccaacagcacttgggttttctctgttgaaggggagctagaagac 300
Db 1636 catcagccaactccccaacagcacttgggttttctctgttgaaggggagctagaagac 1695
Qy 301 aggaactagctggaatttctctgagccaagaaatcccttaagctagctggaggtgact 360
Db 1696 aggaactagctggaatttctctgagccaagaaatcccttaagctagctggaggtgact 1755
Qy 361 gcatccacacttcaaaatgaggtgcttgaacttagctacaccccaacccaacagagagctc 420
Db 1756 gcatccacacttcaaaatgaggtgcttgaacttagctacaccccaacccaacagagagctc 1815
Qy 421 actaaatgctaattagcgaataaagaggttaagaagaatagccaatcatctatgctg 480
Db 1816 actaaatgctaattagcgaataaagaggttaagaagaatagccaatcatctatgctg 1875
Qy 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
Db 1876 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 1935
Qy 541 caaccccttgggtccccccttctgtatcggtctgttttctactatctcactc 600
Db 1936 caaccccttgggtccccccttctgtatcggtctgttttctactatctcactc 1995
Qy 601 attaatcttgcactgcaactgcaaaaaaa 635
Db 1996 attaatcttgcactgcaactgcaaaaaaa 2030

RESULT 4
AAVA3219
ID AAV43219 standard; cDNA: 1329 BP.
XX
XX AAV43219;
XX
XX 29-DEC-1998 (first entry)
XX
XX Multiple sclerosis associated retrovirus fragment 6.
XX
XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
XX gag gene; env gene; rheumatoid arthritis-associated virus; ss.
XX
XX Multiple sclerosis associated retrovirus.
XX

```

Key Location/Qualifiers
 CDS 2..490
 /tag= a
 /product= "Encodes protein AAW71069"
 /transl_except= (pos:177-79, appears to code for a stop codon)
 /transl_except= (pos:125-127, appears to code for a stop codon)
 /transl_except= (pos:137-139, appears to code for a stop codon)
 WO9823755-A1.
 04-JUN-1998.
 26-NOV-1997; 97WO-1B0148Z.
 26-NOV-1996; 96US-0756429.
 (INMR) BIO MERIEUX.
 Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F, Mandrand B, Paranhos-Baccala G, Perron H;
 WPI: 1998-322732/28.
 P-PSDB: AAW71069.
 New nucleic acid from retroviruses - useful for diagnosis, prevention and treatment of, e.g. multiple sclerosis
 Disclosure: Pages 187-188; 286pp: English.
 The present sequence represents a multiple sclerosis (MS) associated retrovirus (MSRV) genomic fragment used in the method of the invention. The invention provides complete or partial genomic sequences of the MSRV-1 pol. gene, gag gene and env gene, and polyepitides encoded by these genes. The invention also provides antibodies raised against the polyepitides. The genomic sequences, polyepitides and antibodies are also claimed useful for diagnosing infection by MS and rheumatoid arthritis-associated viruses, and also for prevention and treatment of infection with these viruses.
 Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other:

Query Match 85.8%; Score 545; DB 19; Length 1329;
 Best Local Similarity 92.4%; Pred. No. 3.3e-159;
 Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

1 ccctgtatcttaacctcctgttaagtgtgtctctccagaatacaaatgtaacta 60
 ||||||| || ||||||| ||||||| ||||||| ||||||| ||
 Db 257 ccctgtatcttaacctcctgttaagtgtgtctctccagaatacaaatgtaacta 316
 Oy 61 caaatgtcttcaaatgtgagcagcagatgagatcagatcagatcagatcagatc 120
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 317 caaaatgtcttcaaatgtgagcagcagatgagatcagatcagatcagatcagatc 376
 Oy 121 ctgagccgagcctgtagcagcagatgagatcagatcagatcagatcagatcagatc 180
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 377 ctgagccgagcctgtagcagcagatgagatcagatcagatcagatcagatcagatc 436
 Oy 181 gaaatctcaactgcaaacccctactatgtcccaattcagcggaggaagcagtagagcggt 240
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 437 gaaatctcaactgcaaacccctactatcaactcaatcagtagaggaagcagtagagcggt 496
 Oy 241 catagcgaacccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 497 tgcagcgaacccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 556
 Oy 301 aggaactagcagatctcctagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 557 aggaactagcagatctcctagcagcagcagcagcagcagcagcagcagcagcagcagcagc 616

Oy 361 gcatccacccttaaacatgagcgttgcaactagctacaccgccagcaatagagagctc 420
 ||||||| || ||||||| ||||||| ||||||| ||||||| ||
 Db 617 gcatccacccttaaacatgagcgttgcaactagctacaccgccagcaatagagagctc 676
 Oy 421 actaaatgcaatattagcagcaaaaatagaggttaagaaatgccaatattatctgctc 480
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 677 actaaatgcaatattagcagcaaaaatagaggttaagaaatgccaatattatctgctc 736
 Oy 481 agagcagcagcggaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 540
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 737 agagcagcagcggaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 796
 Oy 541 caaccctcttggtggtccctcccttgatagggcgtctgttcaactatctcaactct 600
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Db 797 caaccctcttggtggtccctcccttgatagggcgtctgttcaactatctcaactct 856
 Oy 601 attaatcttgcaactgaa 619
 ||||||| ||||||| ||||||| ||
 Db 857 attaatcttgcaactgaa 875

RESULT 5

AAx29704
 ID AAX29704 standard; DNA; 1329 BP.

AAx29704;
 AC AAX29704;

DT 08-JUN-1999 (first entry)
 XX

DE Clone 5M6 from MSRV-1.
 XX

KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 XX

OS rheumatoid polyarthritits; ss.
 XX

PN Multiple sclerosis related virus type 1.
 XX

FR2765588-A1.
 XX

PD 08-JAN-1999.
 XX

PF 07-JUL-1997; 97FR-0008816.
 XX

PR 07-JUL-1997; 97FR-0008816.
 XX

PS (INMR) BIO MERIEUX.
 XX

PA WPI: 1999-098275/09.
 XX

PT P-PSDB: AAW99554.
 XX

PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 XX

PS Claim 1; Page 39-40; 83pp: French.
 XX

CC This sequence represents clone 5M6 from a novel multiple sclerosis
 CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
 CC prophylactic or therapeutic compositions to inhibit expression of a
 CC multiple sclerosis related virus and/or virus associated with
 CC rheumatoid polyarthritits.
 XX

SO Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.8%; Score 545; DB 20; Length 1329;
 Best Local Similarity 92.4%; Pred. No. 3.3e-159;
 Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Oy 1 ccctgtatcttaacctcctgttaagtgtgtctctccagaatacaaatgtaacta 60
 ||||||| || ||||||| ||||||| ||||||| ||||||| ||
 Db 257 ccctgtatcttaacctcctgttaagtgtgtctctccagaatacaaatgtaacta 316
 ||||||| ||||||| ||||||| ||||||| ||||||| ||
 Oy 61 caaatgtcttcaaatgtgagcagcagatgagatcagatcagatcagatcagatcagatc 120
 ||||||| ||||||| ||||||| ||||||| ||||||| ||

Db 317 caaatagttcttcaaatggaaccccgatgacgtacatgactaaactacacgtgagacc 376
QY 121 ctgaccgagcgtctgacccatgctccgagtgttaatgacatgaaagcaccctccgag 180
Db 377 ctgaccgagcgtctgacccatgctccgagtgttaatgacatgaaagcaccctccgag 436
QY 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgggaagcagttagacggt 240
Db 437 gaatctcaactgacaaacccctactactcaactcaattcagtaggaagcagttagacgag 436
QY 241 catcagcaaacctcccaaaagacacttgggtttctgttgaagaggagcagtagagac 300
Db 497 tgcagcacaacccctcccaaaagacacttgggtttctgttgaagaggagcagtagagac 556
QY 301 agactagctggatttccctagtcgcaacgaagaatccctcaactagctgggaaggtgact 360
Db 557 agactagctggatttccctagtcgcaacgaagaatccctcaactagctgggaaggtgact 616
QY 361 gcatcaccctctaacaatggggttgcaacttagctacaccccgaccaatcagaagctc 420
Db 617 gcatcaccctctaacaatggggttgcaacttagctacaccccgaccaatcagaagctc 676
QY 421 actaaatgtctaattgagcaaaaaataggaggttaagaagaatcgaatcatctattgctg 480
Db 677 actaaatgtctaattgagcaaaaaataggaggttaagaagaatcgaatcatctattgctg 736
QY 481 agagcacagcggagagagacaaagatccggatataaaccccgacatcgagccgacaag 540
Db 737 agagcacagcggagagagacaaagatccggatataaaccccgacatcgagccgacaag 796
QY 541 caacccctctgggtcccccctctgttgaaggcgctctgtttcactcatctact 600
Db 797 caacccctctgggtcccccctctgttgaaggcgctctgtttcactcatctact 856
QY 601 attaaatctgcaactgaa 619
Db 857 attaaatctgcaactgaa 875

RESULT 6
AAX77526
ID AAX77526 standard; cDNA; 2946 BP.
XX
XX AAX77526;
XX
DT 10-AUG-1999 (first entry)
XX
DE Human secreted protein A1772-2 cDNA.
XX
M Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
M bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KM cell proliferation; cell differentiation; suppressor; tumour inhibition;
KM hematopoiesis regulator; activin; inhibin; chemotactic; chemokine;
KM haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KM cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
XX Homo sapiens.
OS
PN W09926972-A1.
XX
PD 03-JUN-1999.
XX
PE 17-NOV-1998; 98MO-US24614.
XX
PR 20-OCT-1998; 98US-0175928.
PR 21-NOV-1997; 97US-0976110.
PR 18-MAY-1998; 98US-0080478.
XX
XX (GEMV) GENETICS INST INC.
PA
PI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Merberg D, Treacy M;
XX

DR WPI: 1999-357813/30.
DR P-PSDB: AAY08622.
XX
PT New polynucleotides encoding secreted proteins
XX
PS Claim 13a; Page 100-101; 142pp; English.
XX
XX This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other:

Query Match 84.1%; Score 533.8; DB 20: Length 2946;
Best Local Similarity 91.5%; Pred. NO. 1.5e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

QY 1 cccgtatctcttaacctctctgttaagtgtgtctctccagaatcaaaactgtaaacta 60
Db 223 cccgtatctcttaacctctctgttaagtgtgtctctccagaatcgaaagctgtgaaacta 2382
QY 61 caaatgtcttcaaatgagacacagatgagtgatccatgactaagatccacgttgacc 120
Db 2383 -----caaatgagacacagatgagtgatccatgactaagatccacgttgacc 2430
QY 121 ctgaccgagcgtctagcccatgctccgagtgttaatgacatgaaagcaccctccgag 180
Db 2431 ctgaccgagcgtctagcccatgctccgagtgttaatgacatgaaagcaccctccgag 2490
QY 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgggaagcagttagacggt 240
Db 2491 gaaatctcaactgacaaacccctactactagcccgaatcagcgggaagcagttagacggt 2550
QY 241 catcagcaaacctcccaaaagacacttgggtttctgttgaagaggagcagtagagac 300
Db 2551 cctcgcaaacctcccaaaagacacttgggtttctgttgaagaggagcagtagagac 2610
QY 301 agactagctggatttccctagtcgcaacgaagaatccctcaactagctgggaaggtgact 360
Db 2611 agactagctggatttccctagtcgcaacgaagaatccctcaactagctgggaaggtgact 2670
QY 361 gcatcaccctctaacaatggggttgcaacttagctacaccccgaccaatcagaagctc 420
Db 2671 acatcaccctctaacaacggggttgcaacttagctacaccccgaccaatcagaagctc 2730
QY 421 actaaatgtctaattgagcaaaaaataggaggttaagaagaatcgaatcatctattgctg 480
Db 2731 actaaatgtctaattgagcaaaaaataggaggttaagaagaatcgaatcatctattgctg 2790
QY 481 agagcacagcggagagagacagatcggtgatataaacccggacttgagccggacaag 540
Db 2791 agagcacagcggagagagacagatcggtgatataaacccggacttgagccggacaag 2850
QY 541 caacccctctgggtcccccctctgttgaaggcgctgtttcactcatctact 600
Db 2851 caacccctctgggtcccccctctgttgaaggcgctgtttcactcatctact 2910
QY 601 attaaatctgcaactgaaaaaataaaaaa 635
Db 2911 attaaatctgcaactgaaaaaataaaaaa 2945

```
RESULT 7
AAZ59468
ID AAZ59468 standard; cDNA: 2946 BP.
XX
AC AAZ59468;
XX
DT 11-APR-2000 (first entry)
XX
DE Human secreted protein AJ172_2 polynucleotide sequence.
XX
KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; haematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
MO9960020-A1.
XX
PD 25-NOV-1999.
XX
PE 17-MAY-1999; 99WO-US10915.
XX
PR 18-MAY-1998; 98US-0080478.
XX
PR 20-OCT-1998; 98US-0175928.
XX
PA (GENVY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallic ER, Collins-Racle LA, Evans C;
PI Meckberg D, M1 S, Treacy M;
XX
DR MPI: 2000-116311/10.
XX
P-PSDB; AAY67313.
XX
PT New polynucleotides encoding secreted cDNA libraries, used to develop
PT products for the diagnosis and treatment of neoplastic disease
XX
PS Claim 14; Page 107-108; 149pp; English.
XX
CC This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC from a human adult testes cDNA library. The invention relates to secreted
CC human and murine proteins. The polynucleotides and proteins are predicted
CC to have biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals. Detection of the levels of the proteins can be used for the
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC which modulate the expression or function of the proteins may be used for
CC treating a neoplastic disease and inhibiting metastasis. Other suggested
CC activities include nutritional activity (e.g. in feeds), cytokine and
CC cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The
CC polynucleotide sequences are also stated to be useful for gene therapy.
XX
SO Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 84.1%; Score 533.8; DB 21; Length 2946;
Best Local Similarity 91.5%; Pred. No. 1.5e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;
```

```
Db 2383 -----caaatgagcccaagatgcagtcacagactaagaatctaccgagacc 2420
Qy 121 ctgagccgctgtacgccaatgctccgaatgtaatgacatgaaggaccctcccgag 180
Db 2431 ctgagccgctgtacgccaatgctccgaatgtaatgacatgaaggaccctcccgag 2430
Qy 181 gaaatctcaactgcacaacccctactatgcccgaatctcagcggaagcagttagaagcgt 240
Db 2491 gaaatctcaactgcacaacccctactatgcccgaatctcagcggaagcagttagaagcgt 2250
Qy 241 catcagcaactcccccacagcactgtgttctctgttgaaggcgagctgaagagac 300
Db 2551 cgtcgccaacctcccccacagcactgtgttctctgttgaaggcgagctgaagagac 2610
Qy 301 aggaactagctgattctcctaagcacaagaaatccctaaagcctagctggaaggtgact 360
Db 2611 aggaactagctgattctcctaagcacaagaaatccctaaagcctagctggaaggtgact 2670
Qy 361 gcatccacccttaaacatgaggctgcaacttagctcacaccgccaatcagaagctc 420
Db 2671 acatccacccttaaacacagcggtgtgcaacttagctcacaccgccaatcagaagctc 2730
Qy 421 actaaatgctcaattaggcaaaaatagaggttaagaatgccaatcattatctgctg 480
Db 2731 actaaatgctcaattaggcaaaaatagaggttaagaatgccaatcattatctgctg 2790
Qy 481 agagcacagcgagaggaagcaagatcggagatataaacccaagcaltcagcgcaagcg 540
Db 2791 agagcacagcgagaggaagcaagatcggagatataaacccaagcaltcagcgcaagcg 2850
Qy 541 caacccttgggtgccctcttggatgagcgctgtttcactctatcactc 600
Db 2851 caacccttgggtgccctcttggatgagcgctgtttcactctatcactc 2910
Qy 601 attaatctgcaactgaaaaaataaaaaa 635
Db 2911 attaatctgcaactgcaaaaaaataaaaaa 2945

RESULT 8
AAF55630
ID AAF55630 standard; DNA: 2781 BP.
XX
AC AAF55630;
XX
DT 29-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FH CDS 762..2378
FH FT /*tag= a
FH FT /product= "envelope protein"
XX
PN WO200116171-A1.
XX
PD 08-MAR-2001.
XX
PE 01-SEP-2000; 2000WO-FR02429.
XX
PR 01-SEP-1999; 99FR-0011141.
XX
PR 15-SEP-1999; 99FR-0011793.
XX
PA (INMR ) BIO MERIEUX.
XX
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI Mallet F, Cosset F, Blond J, Lavalliette D, Bouton O, Ruggieri A;
```

DR WPI; 2001-226676/23.
P-PSDB; AAB67652.

XX Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia -
XX

PS Disclosure; Page 44-45; 57pp; French.

XX The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV). In
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.

Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other:

Query Match 83.8%; Score 532.2; DB 22; Length 2781;
Best Local Similarity 91.3%; Pred. No. 4.5e-155;
Matches 580; Conservative 0; Mismatches 43; Indels 12; Gaps 1;

```

QY 1 cccgtatctttaactcctctgttaagtgtctctccagaatcaaaactgtaaacta 60
DB 2157 cccgtatctttaactcctctgttaagtgtctctccagaatcaaaactgtaaacta 2216
QY 61 caattgtctcaaatgtgacccagatgagtcagatcagatcccccgtgagcc 120
DB 2217 -----caattgtgacccagatgagtcagatcagatcccccgtgagcc 2264
QY 121 ctgagccggtctgtagcccatgtccgagttaatgacatgaaagccaccctccgag 180
DB 2265 ctgagccggtctgtagcccatgtccgagttaatgacatgaaagccaccctccgag 2324
QY 181 gaaatcacaatgacaaacccctactatgcccgaatcagcgggaagcagtagacggt 240
DB 2325 gaaatcagcgtgacaaacccctactatgcccgaatcagcgggaagcagtagacggt 2384
QY 241 catcagcaaacctcccaacagcactgtggtttccgttggagaggggagcagaagac 300
DB 2385 cgtcggcaaacctcccaacagcactgtggtttccgttggagaggggagcagaagac 2444
QY 301 aggaactagctgatttctctagagccaacgaagaatccctaaagcctagctgggaagtgact 360
DB 2445 aggaactagctgatttctctagagccaacgaagaatccctaaagcctagctgggaagtgact 2504
QY 361 gcatcaactctaaacatggtggtctgcaacttagctcacaccgaccaatcagaagctc 420
DB 2505 acatcaactctaaacatggtggtctgcaacttagctcacaccgaccaatcagaagctc 2564
QY 421 actaaatgttaattgagcaaaaataggaagttaaagaatagccaatcatctttgctgt 480
DB 2565 actaaatgttaattgagcaaaaataggaagttaaagaatagccaatcatctttgctgt 2624
QY 481 agagcagcagggagggagcaagaatcggaatataaacccagcatctcagcggcagaag 540
DB 2625 agagcagcagggagggagcaagaatcggaatataaacccagcatctcagcggcagaag 2684
QY 541 caaccctcttggtccctcctcttgataggcgctctgttttcaactcatctactctc 600
DB 2685 caaccctcttggtccctcctcttgataggcgctctgttttcaactcatctactctc 2744
QY 601 attaatcttgcaactgaaataaaaaaa 635
DB 2745 attaatcttgcaactgcaataaaaaaa 2779

```

RESULT 9
ID AAX25660
AC AAX25660; standard; cDNA to mRNA; 1136 BP.

XX AAX25660;
AC AAX25660;
XX 21-MAY-1999 (first entry)

DE Human endogenous retrovirus W clone cl.C4C5.

XX Clone: human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX

OS Human endogenous retrovirus.

PN W09902696-A1.

PD 21-JAN-1999.

PF 06-JUL-1998; 98WO-FR01442.

PR 07-JUL-1997; 97FR-0008815.

PA (INMR) BIO MERIEUX.

PI Besene F, Blond JL, Boulton O, Mallet F, Mandrand B.

DR WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
PS Claim 1; Page 59-60; 106pp; French.

XX This sequence represents clone cl.C4C5 of the human endogenous retrovirus
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.

Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other:

Query Match 82.8%; Score 525.8; DB 20; Length 1136;
Best Local Similarity 90.7%; Pred. No. 2.9e-153;
Matches 576; Conservative 0; Mismatches 47; Indels 12; Gaps 1;

```

QY 1 cccgtatctttaactcctctgttaagtgtctctccagaatcaaaactgtaaacta 60
DB 510 cccgtatctttaactcctctgttaagtgtctctccagaatcaaaactgtaaacta 569
QY 61 caattgtctcaaatgtgacccagatgagtcagatcagatcccccgtgagcc 120
DB 570 -----caattgtgacccagatgagtcagatcagatcccccgtgagcc 617
QY 121 ctgagccggtctgtagcccatgtccgagttaatgacatgaaagccaccctccgag 180
DB 618 ctgagccggtctgtagcccatgtctgtagatgacatcaaaagccaccctctgag 677
QY 181 gaaatcacaatgacaaacccctactatgcccgaatcagcgggaagcagtagagcgt 240
DB 678 gaaatcagcgtgacaaacccctactatgcccgaatcagcgggaagcagtagagcgt 737
QY 241 catcagcaaacctcccaacagcactgtggtttccgttggagaggggagcagaagac 300
DB 738 cgtcggcaaacctcccaacagcactgtggtttccgttggagaggggagcagaagac 797
QY 301 aggaactagctgatttctctagagccaacgaagaatccctaaagcctagctgggaagtgact 360

```



```
Db 798 aggaatagctggtattcttctagctgactaagaatccccaagcctcggaggaagtgcac 857
Oy 361 gcatccacctcaaaacaggggctgtcaactagctcaccccgacacatcagaagctc 420
Db 858 acatccaccttaaaacaggggctgtcaactagctcaccccgacacatcagaagctc 917
Oy 421 actaaatgctaatatagcaaaaabaggaagtaagaataagccaatcatctatgctg 480
Db 918 actaaatgctaatatagcaaaaabaggaagtaagaataagccaatcatctatgctg 977
Oy 481 agagacacgagggagggaggaagatcgatataaaccagcatctgagccggagacg 540
Db 978 agagacacgagggagggaggaagatcgatataaaccagcatctgagccggagacg 1037
Oy 541 caaccctctgggtccctccctctgtatgagcgctctgttccactctatctactc 600
Db 1038 caaccctctgggtccctccctctgtatgagcgctctgttccactctatctactc 1097
Db 601 attaaatctgcaactgaaaaaagaaaaa 635
1098 attaaatctgcaactgaaaaaagaaaaa 1132
```

RESULT 10

AAAS9210 standard; DNA; 1136 BP.

AAAS9210;

07-NOV-2000 (first entry)

3' pol gene and 3' non coding sequences of HERV-W from human genome.

Autoimmune disease: retrovirus: human endogenous retrovirus W: HERV-W; gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

Homo sapiens.

MO200043521-A2.

27-JUL-2000.

21-JAN-2000; 2000WO-FR00144.

21-JAN-1999; 99FR-0000888.

(IMMR) BIO MERIEUX.

Paranhos-Baccala G, Mallet F, Volsset C;

WPI: 2000-499229/44.

New nucleic acid from human endogenous retrovirus: useful e.g. for diagnosis of autoimmune disease and complications of pregnancy, contains at least part of the gag gene

Disclosure: Page 46; 53pp; French.

The present sequence represents an endogenous retroviral nucleic acid fragment, which is associated with an autoimmune disease, and is integrated into the human genome. The fragment is originally derived from a novel retrovirus, human endogenous retrovirus W (HERV-W). The HENV-W retrovirus is associated with autoimmune disease, failure of pregnancy or disorders of pregnancy. The nucleic acid fragment, or proteins derived from it, are useful for diagnosis of autoimmune disease (specifically multiple sclerosis) and for monitoring pregnancy. The nucleic acid fragments may also be used for in situ labelling of isolated chromosomes, while the transcription product can be used to study or monitor T cell proliferation in vitro.

Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other;

Query Match 82.8%; Score 525.8; DB 21; Length 1136;
Best Local Similarity 90.7%; Pred. No. 2.9e-153;
Matches 576; Conservative 0; Mismatches 47; Indels 12; Gaps 1;

```
Oy 1 cccgtatcttaactctctgttaagtgtctctctccaaatcaaacgttaacta 60
Db 510 cccgtatcttaactctctgttaactgtctctctccaaatcaaacgttaacta 569
Oy 61 caattgtcttcaaatgagcacagaatgagatccaaatgaatcaccgagacc 120
Db 570 -----caattgagaccaaagtgcagtcacaaatgaatcaccgagacc 617
Oy 121 ctgaccgacctgtacgccaatgctccgaatgtaatgacatgagcaccctccgag 180
Db 618 ctgaccgacctgtacgccaatgctccgaatgtaatgacatgagcaccctccgag 677
Oy 181 gaatctcaatgtcacaacccctctatgcccacatcagaagggaagttaagcgt 240
Db 678 gaatctcaatgtcacaacccctctatgcccacatcagaagggaagttaagcgt 737
Oy 241 catcagccaacctccccaacagcaactgtggttccgtgtgagaggggagctgagac 300
Db 738 cgtcgccaacctccccaacagcaactgtggttccgtgtgagagggagctgagac 797
Oy 301 aggaactagctgattctcctagggccaacgaagaatccctaagcctgaggagtgact 360
Db 798 aggaactagctgattctcctagggccaacgaagaatccctaagcctgaggagtgact 857
Oy 361 gcatccaccttaaaacaggggctgtgcaactagctcaacacccgacatcagaagctc 420
Db 858 acatccacctttaaacaggggctgtgcaactagctcaacacccgacatcagaagctc 917
Oy 421 actaaatgctaatatagcaaaaabaggaagtaagaataagccaatcatctatgctg 480
Db 918 actaaatgctaatatagcaaaaabaggaagtaagaataagccaatcatctatgctg 977
Oy 481 agagacacgagggagggaggaagatcgatataaaccagcatctgagccggagacg 540
Db 978 agagacacgagggagggaggaagatcgatataaaccagcatctgagccggagacg 1037
Oy 541 caaccctctgggtccctccctctgtatgagcgctctgttccactctatctactc 600
Db 1038 caaccctctgggtccctccctctgtatgagcgctctgttccactctatctactc 1097
Oy 601 attaaatctgcaactgaaaaaagaaaaa 635
Db 1098 attaaatctgcaactgaaaaaagaaaaa 1132
```

RESULT 11

AAAX25661 standard; cDNA to mRNA; 2782 BP.

AAAX25661;

21-MAY-1999 (first entry)

Human endogenous retrovirus W clone cl. PH74.

Clone: human endogenous retrovirus; genome: autoimmune disease; multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes; disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

Human endogenous retrovirus.

W09902696-A1.

21-JAN-1999.

06-JUL-1998; 98WO-FR01442.

07-JUL-1997; 97FR-0000815.

XX (INMR) BIO MERIEUX.
 PA Beseme F, ~ Blond JL, Bouton O, Mallet F, Mandrand B;
 PI WPI: 1999-120897/10.
 XX
 XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy
 XX
 XX Claim 1: Page 60-63; 106pp; French.
 XX
 CC This sequence represents clone cl.PH74 of the human endogenous retrovirus
 CC (HERV) W genome. The nucleic acid, their fragments or peptides encoded
 CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
 CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
 CC dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.

Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 82.3%; Score 522.6; DB 20; Length 2782;
 Best Local Similarity 90.4%; Pred. No. 4,4e-152;

Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

QY 1 ccctgtatctttaaactcctgttaagtgtgtctctccagaatcaaaactgtanaacta 60
 DB 2158 cccgtatctttaaactcctgttaagtgtgtctctccagaatcaaaactgtanaacta 2217
 QY 61 caattgttcttaaatgagccacagatgagttcattagatccaccctgtgagcc 120
 DB 2218 -----caatgagccacagatgagttcattagatccaccctgtgagcc 2265
 QY 121 ctgagccgctgtctgacccatgctccgattgattacattgaagccaccctccgag 180
 DB 2266 ctgagccgctgtctgacccatgctccgattgattacattgaagccaccctccgag 2325
 QY 181 gaaatcctaactgacaaacccactactatgcccgaatcagcgggaagcagtagagcgt 240
 DB 2326 gaaatcctaactgacaaacccactactatgcccgaatcagcgggaagcagtagagcgt 2385
 QY 241 catcagcacaactccccaagaactgttttctgttgaagagggagcagtagagac 300
 DB 2386 ggtcggcacaactccccaagaactgttttctgttgaagagggagcagtagagac 2445
 QY 301 agagctagctgatttctcctagcacaagaatccctaagctagctggaaggtgac 360
 DB 2446 agagctagctgatttctcctagcacaagaatccctaagctagctggaaggtgac 2505
 QY 361 gcatcacaacttaaacatgaggctgtgcaactagctcacaccgacacacagagagctc 420
 DB 2506 acatccacactttaaacaacgggctgtgcaactagctcacaccgacacacagagagctc 2565
 QY 421 actaaatgttaatttggaataatggaagttaagaataagccaatcattatgtgctg 480
 DB 2566 actaaatgttaatttggaataatggaagttaagaataagccaatcattatgtgctg 2625
 QY 481 agagcacaagcagagagacagatcgagatatcaaccgcatctgagccggcagcgg 540
 DB 2626 agagcacaagcagagagacagatcgagatatcaaccgcatctgagccggcagcgg 2665
 QY 541 caaccctcttggtccctcccttctgtatggcgtctgttttactactatctactct 600
 DB 2686 caaccctcttggtccctcccttctgtatggcgtctgttttactactatctactct 2745
 QY 601 attaatctgtcactgtaaaaaaaatgagagagagagagagagagagagagagagag 635
 DB 2746 attaatctgtcactgtaaaaaaaatgagagagagagagagagagagagagagagag 2780

RESULT 12

AAA59211 standard; DNA; 2782 BP.

AAA59211;

07-NOV-2000 (first entry)

5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.

Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W; gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

Homo sapiens.

WO200043521-A2.

27-JUL-2000.

21-JAN-2000; 2000MO-FR00144.

21-JAN-1999; 99FR-0000888.

(INMR) BIO MERIEUX.

Paranhos-Baccala G, Mallet F, Voisset C;

WPI: 2000-499229/44.

New nucleic acid from human endogenous retrovirus, useful e.g. for diagnosis of autoimmune disease and complications of pregnancy, contains at least part of the gag gene

Disclosure; Page 46-47; 53pp; French.

The present sequence represents an endogenous retroviral nucleic acid fragment, which is associated with an autoimmune disease, and is integrated into the human genome. The fragment is originally derived from a novel retrovirus, human endogenous retrovirus W (HERV-W). The HERV-W retrovirus is associated with autoimmune disease, failure of pregnancy or disorders of pregnancy. The nucleic acid fragment, or proteins derived from it, are useful for diagnosis of autoimmune disease (specifically multiple sclerosis) and for monitoring pregnancy. The nucleic acid fragments may also be used for in situ labelling of isolated chromosomes, while the transcription product can be used to study or monitor T cell proliferation in vitro.

Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 82.3%; Score 522.6; DB 21; Length 2782;
 Best Local Similarity 90.4%; Pred. No. 4,4e-152;

Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

QY 1 ccctgtatctttaaactcctgttaagtgtgtctctccagaatcaaaactgtanaacta 60
 DB 2158 cccgtatctttaaactcctgttaagtgtgtctctccagaatcaaaactgtanaacta 2217
 QY 61 caattgttcttaaatgagccacagatgagttcattagatccaccctgtgagcc 120
 DB 2218 -----caatgagccacagatgagttcattagatccaccctgtgagcc 2265
 QY 121 ctgagccgctgtctgacccatgctccgattgattacattgaagccaccctccgag 180
 DB 2266 ctgagccgctgtctgacccatgctccgattgattacattgaagccaccctccgag 2325
 QY 181 gaaatcctaactgacaaacccactactatgcccgaatcagcgggaagcagtagagcgt 240
 DB 2326 gaaatcctaactgacaaacccactactatgcccgaatcagcgggaagcagtagagcgt 2385
 QY 241 catcagcacaactccccaagaactgttttctgttgaagagggagcagtagagac 300
 DB 2746 attaatctgtcactgtaaaaaaaatgagagagagagagagagagagagagagagag 2780

```

Db 2386 gtcgccaacctcccaacagcacttaggtttctctgttagatgagggaagagac 2445
Oy 301 aggaactagctgatttctctagagcaacgaagaatccctaaagctagcgggaagtgact 360
    |||||||
Db 2446 aggaactagctgatttctctagagcaacgaagaatccctaaagctagcgggaagtgact 2505
Oy 361 gcatccacctctaaacatgagggtctgcaacttagctcacaccgcgaacatcagaagagctc 420
    |||||||
Db 2506 acatccacctctaaacagcgggtctgcaacttagctcacaccgcgaacatcagaagagctc 2565
Oy 421 actaaatgcttaattagcaaaaataagaggttaagaatagccatcatctatgtgctg 480
    |||||||
Db 2566 actaaatgcttaattagcaaaaataagaggttaagaatagccatcatctatgtgctg 2625
Oy 481 agagcagcagcgaggaagcaagatcggagataaaacccagcattcgagccgcgaacg 540
    |||||||
Db 2626 agagcagcagcgaggaagcaagatcggagataaaacccagcattcgagccgcgaacg 2685
Oy 541 caaccctcttggtgctccctcccttgatgaggcgtctgttttcaacctattcaactct 600
    |||||||
    2666 caaccctcttggtgctccctcccttgatgaggcgtctgttttcaacctattcaactct 2745
Oy 601 attaatctgcaactgaaaaaataaaaaa 635
    |||||||
Db 2746 attaatctgcaactgaaaaaataaaaaa 2780

RESULT 13
AAH20069
ID AAH20069 standard; DNA: 2782 BP.
AC
XX AAH20069:
XX
XX 08-AUG-2001 (first entry)
XX
DE HERV-W envelope protein G encoding nucleic acid.
XX
KW Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW surface antigen; multiple sclerosis-related superantigen; vaccine;
KW antisense-therapy; autoimmune disorder; ds.
XX
OS Human endogenous retrovirus.
XX
XX
XX Key Location/Qualifiers
XX 5'UTR 1..762
XX CDS 763..2379
    /*tag= a
    /*tag= b
    /*product= "HERV-W envelope protein G"
    /transl_except= (pos:790..792,aa:Phe)
    /transl_except= (pos:793..795,aa:Thr)
    /transl_except= (pos:812..814,aa:Leu)
    /transl_except= (pos:818..820,aa:Ser)
    /transl_except= (pos:862..864,aa:Tyr)
    /transl_except= (pos:865..867,aa:Gln)
    /transl_except= (pos:1174..1176,aa:Arg)
    /transl_except= (pos:1441..1443,aa:Leu)
    /transl_except= (pos:1903..1905,aa:Gln)
    /transl_except= (pos:2017..2019,aa:Lys)
    /transl_except= (pos:2026..2028,aa:Arg)
    /transl_except= (pos:2044..2046,aa:Arg)
    /transl_except= (pos:2089..2091,aa:Gln)
    /transl_except= (pos:2170..2172,aa:Asn)
    2380..2782
    /*tag= c
XX
XX WO200131021-A1.
XX
XX 03-MAY-2001.
XX
XX 30-OCT-2000; 2000WO-EPI0659.
XX

```

```

PR 28-OCT-1999; 99EP-0402690.
PA
XX (UYGE-) UNIV GENEVE.
XX
XX Conrad B, Mach B;
XX
XX WPI: 2001-316336/33.
XX P-PSDB; AAB75138.
XX
XX New human retrovirus HERV-W ENV proteins/peptides having superantigen
XX activity useful for diagnosing and treating multiple sclerosis -
XX
XX Claim 13; Fig 9; 94pp; English.
XX
XX On the basis of the PBS t-RNA motif used for the classification of human
XX endogenous retrovirus (HERVs) the full length endogenous provirus which
XX has been located on the long arm of human chromosome 7 (7q21-22) has been
XX designated HERV-W. The present invention describes proteins or peptides
XX (I) having superantigen (Sag) activity comprising the ENV protein (ENV)
XX of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
XX have neuroprotective activity, and can be used in: vaccines; antisense-
XX therapy; and HERV-W Sag activity-inhibitors. (I) and encoding DNA/RNA are
XX useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
XX disorders. (I) are also useful for identifying substances (and optionally
XX recovering) capable of binding to a retroviral superantigen associated
XX with MS, substances capable of blocking Sag activity and substances
XX capable of blocking transcription or translation of HERV-W retroviral
XX superantigen. A protein or peptide derived from (I), modified to be
XX devoid of Sag activity and being capable of generating an immune response
XX against HERV-W retroviral Sag is useful in therapy. Nucleic acid
XX molecules encoding (I) are useful as vaccines against MS. Substances
XX capable of blocking Sag activity, capable of binding to a retroviral
XX superantigen associated with MS, or capable of blocking transcription or
XX translation of HERV-W retroviral superantigen for use in treating or
XX preventing MS, obtained using (I) are useful for the treatment and
XX prevention of MS. (I) and nucleic acids encoding them are useful for
XX diagnosing autoimmune disease. The present sequence encodes the
XX specifically claimed envelope protein of HERV-W designated G.
XX
XX Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 other:

Query Match 82.3%; Score 522.6; DB 22; Length 2782;
Best Local Similarity 90.4%; Pred. No. 4,4e-152;
Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

Oy 1 cccgtatcttaaacctctgttaagtgtgtcttcacgaatacaaaacta 60
    |||||||
Db 2158 cccgtatcttaaacctctgttaagtgtgtcttcacgaatacaaaacta 2217
Oy 61 caaatgttctcaaatgagcaccagatgagatcagataagatccaccgtgacc 120
    |||||||
Db 2218 -----caaatgagcccaagatgagatcagataagatccaccgtgacc 2265
Oy 121 ctgagccgctgtcagccatgctccgattatagatcagtaagagccctccgag 180
    |||||||
Db 2266 ctgagccgctgtcagccatgctccgattatagatcagtaagagccctccgag 2325
Oy 181 gaaatcctaactgcaaacccctactatgcccgaattcagcgggaagctagaagcgt 240
    |||||||
Db 2326 gaaatcctaactgcaaacccctactatgcccgaattcagcgggaagctagaagcgt 2385
Oy 241 catcagcaaacctcccaacagcacttggttctctgttgagaggggagctagaagac 300
    |||||||
Db 2386 gtcgccaacctcccaacagcacttggttctctgttgagaggggagctagaagac 2445
Oy 301 aggaactagctgatttctctagagcaacgaagaatccctaaagctagcgggaagtgact 360
    |||||||
Db 2446 aggaactagctgatttctctagagcaacgaagaatccctaaagctagcgggaagtgact 2505
Oy 361 gcatccacctctaaacatgagggtctgcaacttagctcacaccgcgaacatcagaagagctc 420
    |||||||
Db 2506 acatccacctctaaacagcgggtctgcaacttagctcacaccgcgaacatcagaagagctc 2565

```

QY 421 actaaatgctaaatgagcaaaaatagagtaagaagaatcaatcatctatgctg 480
 |||||||
 Db 2566 actaaatgctaaatgagcaaaaatagagtaagaagaatcaatcatctatgctg 2625
 QY 481 agagcacaagcagagagagacagatcgagatataaccagcattcgagccgcaacg 540
 |||||||
 Db 2626 agagcacaagcagagagacagatcgagatataaccagcattcgagccgcaacg 2685
 QY 541 caaccccttggtgctccctcttgatgagcgctctgttttcaatctatctact 600
 |||||||
 Db 2686 caaccccttggtgctccctcttgatgagcgctctgttttcaatctatctact 2745
 QY 601 attaatctgcaactgaaaaaataaaaaa 635
 |||||||
 Db 2746 attaatctgcaactgcaactgcaactgcaactgcaactgcaactgcaactg 2780
 RESULT 14
 AAX25665
 AAX25665 standard: cDNA to mRNA: 7582 BP.
 AC AAX25665;
 DT 21-MAY-1999 (first entry)
 DE Complete human endogenous retrovirus W genome.
 KW Clone: human endogenous retrovirus; genome: autoimmune disease:
 KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.
 XX
 PN W09902696-A1.
 PD 21-JAN-1999.
 PF 06-JUL-1998: 98MO-FR01442.
 PR 07-JUL-1997: 97FR-0008815.
 XX
 PA (INMR) BIO MERIEUX.
 PI Besene F, Blond JL, Bouton O, Mallet F, Mandrand B;
 XX
 DR WPI: 1999-120897/10.
 XX
 PT New nucleic acid sequences from human endogenous retrovirus-W -
 expressed exclusively in placenta and useful in diagnosis and
 therapy of autoimmune disease, and abnormal or failed pregnancy
 Claim 1; Page 71-74; 106pp; French.
 PS This sequence represents the complete sequence of the human endogenous
 CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
 CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
 CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
 CC insulin-dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.
 XX
 SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;
 Query Match 80.5%; Score 511.4; DB 20; Length 7582;
 Best Local Similarity 89.5%; Pred. No. 2.2e-148;
 Matches 552; Conservative 11; Mismatches 42; Indels 12; Gaps 1;
 QY 1 cccggtatcttaaccctcctgttaagttgtctctccagacaataactgttaacta 60
 |||||||
 Db 6976 cccgtatctttraccctcctgttaactgtctctccagacaataactgtttraaacta 7035

QY 61 caaatgttctcaaatgagcaaccagatgagttccatgactaagatccacgtgagcc 120
 |||||||
 Db 7036 -----caaatgagcccaagatgagttccatgactaagatccacgtgagcc 7083
 QY 121 ctgagccgctctgctagaccatgtctccagatgttaatgacatgaaagccctccag 180
 |||||||
 Db 7084 ctgagccgctctgctagaccatgtctccagatgttaatgacatgaaagccctccag 7143
 QY 181 gaaatctcaactgacacccctactatgcccgaatcagcgggaagcagttgagcgt 240
 |||||||
 Db 7144 gaaatctcaactgacacccctactatgcccgaatcagcgggaagcagttgagcgt 7203
 QY 241 catcagcacaaccccccacagacattggttttccgtgtgagaggggagcagagac 300
 : || |||||
 Db 7204 gtcgagcacaaccccccacagacattggttttccgtgtgagaggggagcagagac 7263
 QY 301 aggaactagctgatttccatgacacagaaatccctaaagcctagctggaagtgact 360
 |||||||
 Db 7264 aggaactagctgatttccatgacacagaaatccctaaagcctagctggaagtgact 7323
 QY 361 gcatcacccttaaacatgaggtctgcaacttagctacacccgacacatcagagagctc 420
 |||||||
 Db 7324 acatccacctttaaacaacagcgggtctgcaacttagctacacccgacacatcagagagctc 7383
 QY 421 actaaatgctaaatgagcaaaaatagagagtaagaagaatagccaatcatctatgctg 480
 |||||||
 Db 7384 actaaatgctaaatgagcaaaaatagagagtaagaagaatagccaatcatctatgctg 7443
 QY 481 agagcacaagcagagagacagatcgagatataaccagcattcgagccgcaacg 540
 |||||||
 Db 7444 agagcacaagcagagagacagatcgagatataaccagcattcgagccgcaacg 7503
 QY 541 caaccccttggtgctccctcttgatgagcgctctgttttcaatctatctactc 600
 |||||||
 Db 7504 caaccccttggtgctccctcttgatgagcgctctgttttcaatctatctactc 7563
 QY 601 attaatctgcaactg 617
 |||||||
 Db 7564 attaatctgcaactg 7580
 RESULT 15
 AAA59215
 ID AAA59215 standard: DNA: 7582 BP.
 XX
 AC AAA59215;
 DT 07-NOV-2000 (first entry)
 DE Human endogenous retrovirus W (HERV-W) sequence.
 KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
 KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
 OS Human endogenous retrovirus.
 XX
 FH Key location/Qualifiers
 FT LTR 1..120
 FT /*tag- a
 FT /*note- "R of 5' LTR"
 FT 121..575
 FT /*tag- b
 FT /*note- "U5 of 5' LTR"
 FT 579..596
 FT primer_bind
 FT /*tag- c
 FT 5581..7194
 FT /*tag- d
 FT /*note- "ORF1 env538"
 FT 7039..7194
 FT /*tag- e
 FT /*note- "ORF2 52 AA"
 FT 7112..7255
 FT CDS

```

FT      /tag= f
FT      /note= "ORF3 48 AA"
FT      misc-feature
FT      7244..7254
FT      /tag= g
FT      /note= "polypurine tract"
FT      LTR
FT      7256..7582
FT      /tag= h
FT      /note= "U3-R of 3' LTR"
FT      polyA_signal
FT      7563..7569
FT      /tag= 1
FT      WO20043521-A2.
XX      27-JUL-2000.
XX      21-JAN-2000; 2000MO-FR00144.
XX      21-JAN-1999; 99FR-0000888.
XX      (INMR ) BIO MERIEUX.
XX      Paranhos-Bacala G, Mallet F, Voliset C;
XX      WPI: 2000-499229/44.
XX      New nucleic acid from human endogenous retrovirus, useful e.g. for
XX      diagnosis of autoimmune disease and complications of pregnancy,
XX      contains at least part of the gag gene
XX      PS      Disclosure: Page 49-52; 53pp; French.
XX      The present sequence represents an endogenous retrovirus, which is
XX      associated with an autoimmune disease, and is integrated into the human
XX      genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
XX      HERV-W retrovirus is associated with autoimmune disease, failure of
XX      pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
XX      proteins derived from it, are useful for diagnosis of autoimmune
XX      disease (specifically multiple sclerosis) and for monitoring pregnancy.
XX      The nucleic acid fragments may also be used for in situ labelling of
XX      isolated chromosomes, while the transcription product can be used to
XX      study or monitor T cell proliferation in vitro.
XX      SO      Sequence 7582 BP: 2156 A; 1876 C; 1538 G; 1796 T; 216 other;

```

```

Query Match      80.5%; Score 511.4; DB 21; Length 7582;
Best Local Similarity 89.5%; Pred. No. 2.2e-148;
Matches 552; Conservative 11; Mismatches 42; Indels 12; Gaps 1;
1 cccgtatcttaacctctgttaagttgtctcttccagaatcaaatgtaacta 60
Db      cccgtatctttracctctgttaacttctcttccagaatcgaatgtaacta 7035
OY      61 caaatgtcttcaaatgagacagatgagatcagatcaatcaacgtgagacc 120
Db      7036 -----caaatgagcccaagatgcagccaagaactaactacccgacacc 7083
OY      121 ctgagccgctctagaccatgctccgagttaatgacatgaaagcaccctccag 180
Db      7084 ctgagccgctctgagccagatctgatttaatgacatcaagacaccctctctgag 7143
OY      181 gaaatcgaactgcaaacctctactatgcccgaatlcagcggaagcagttagaagcgt 240
Db      7144 gaaatcgaactgcaaacctctactatgcccgaatlcagcggaagcagttagaagcgt 7203
OY      241 catcgagcaacctcagccagcagacttggtttctctgtgaaggaggagctagaagac 300
Db      7204 sgctggccaacctccccaacacactagttcttctgtgagatcgaggagcagagac 7263
OY      301 aggaactagctgattctcttaggccaagaaatccctaaagcctagctggaagtgact 360
Db      7264 aggaactagctgattctcttaggccaagaaatccctaaagcctagctggaagtgact 7323

```

```

OY      361 gcatccaccttaacaatgaggcttgcaactagctcacaccgaccaatcagagagctc 420
Db      7324 acatccacctttaaacaacgaggcttgcaactagctcacaccgaccaatcagagagctc 7383
OY      421 actaaatgcttaattagagcaaaaatagagggaagaagaatagccaatcattatggcctg 480
Db      7384 actaaatgcttaattagagcaaaaatagagggaagaagaatagccaatcattatggcctg 7443
OY      481 agagcacagcgaggagacaagaagatcgagataaaacccaagcattcgagccgcaacg 540
Db      7444 agagcacagcgaggagacaagaagatcgagataaaacccaagcattcgagccgcaacg 7503
OY      541 caacccttggtggtccctcccttgataggggctgtgttcaactatcaactct 600
Db      7504 caacccttggtggtccctcccttgataggggctgtgttcaactatcaactct 7563
OY      601 attaatcttgactg 617
Db      7564 attaatcttgactg 7580

```

Search completed: June 20, 2002, 09:42:19
Job time: 8469 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 06:44:35 ; Search time 4429.18 Seconds
(without alignments)
3000.183 Million cell updates/sec

Title: US-09-319-156a-6

Perfect score: 635
Sequence: 1 ccctgtattcttaacctctc.....tgaaaaaaaaaaaaaaaaaa 635

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description
------------	--------------------	----	-------------

1	635	100.0	635	6	AX001024	AX001024 Sequence
2	635	100.0	2004	14	AF127229	AF127229 Multiple
3	631.8	99.5	176095	2	AC019346	AC019346 Homo sapi
4	631.8	99.5	210336	2	AC022171	AC022171 Homo sapi
5	580.2	91.4	148724	2	AC010778	AC010778 Homo sapi
6	579.2	91.2	8339	2	AL162912	AL162912 Human DNA
7	571	89.9	136901	9	AC073626	AC073626 Homo sapi
8	564.6	88.9	89728	9	AL583805	AL583805 Human DNA
9	564.6	88.9	212096	2	AL451130	AL451130 Homo sapi
10	560.2	88.2	175040	2	AC072023	AC072023 Homo sapi
11	556.6	87.7	174019	2	AP001538	AP001538 Homo sapi
12	556.6	87.7	340000	2	AP001674	AP001674 Homo sapi
13	551.8	86.9	149755	2	AP001545	AP001545 Homo sapi
14	551.8	86.9	152775	2	AC090313	AC090313 Homo sapi
15	551.8	86.9	163803	2	AC095531	AC095531 Homo sapi
16	551.8	86.9	179109	2	AC107075	AC107075 Homo sapi
17	551.4	86.8	128468	9	AL135090	AL135090 Human DNA
18	551.4	86.8	140756	2	AL135038	AL135038 Human DNA
19	551.4	86.8	167366	2	AC021774	AC021774 Homo sapi
20	551.4	86.8	183499	2	AL607153	AL607153 Homo sapi
21	550.2	86.6	164211	2	AP002790	AP002790 Homo sapi
22	550.2	86.6	168055	2	AC064801	AC064801 Homo sapi
23	549	86.5	180523	2	CNS01DSH	AL121784 Human chr
24	549	86.5	190565	2	AC007374	AC007374 Homo sapi
25	547.6	86.2	192178	2	AC009443	AC009443 Homo sapi
26	547.6	86.2	192899	2	CNS05TF5	AL365295 Human chr
27	547.4	86.2	172281	9	AC068492	AC068492 Homo sapi
28	547.2	86.2	168195	2	AC104163	AC104163 Homo sapi
29	546.6	86.1	158564	2	AC104444	AC104444 Homo sapi
30	546	86.0	169234	2	CNS06C7Q	AL390799 Human chr
31	545.6	85.9	170730	30	AC073868	AC073868 Homo sapi
32	545	85.8	1329	6	AX001030	AX001030 Sequence
33	545	85.8	99408	2	AL359385	AL359385 Human DNA
34	545	85.8	170586	2	AL356126	AL356126 Homo sapi
35	543.8	85.6	197043	2	AC087897	AC087897 Homo sapi
36	543.4	84.9	88328	2	AL357874	AL357874 Human DNA
37	538.8	84.9	122969	2	AC105029	AC105029 Homo sapi
38	537.8	84.7	83412	9	AC092843	AC092843 Homo sapi
39	537.8	84.7	176600	9	AC046130	AC046130 Homo sapi
40	534.4	84.2	175923	2	AC064839	AC064839 Homo sapi
41	534.4	84.2	187289	2	AC083908	AC083908 Homo sapi
42	533.8	84.1	2946	6	ARI17269	ARI17269 Sequence
43	532.2	83.8	2781	6	AX092223	AX092223 Sequence
44	532.2	83.8	2781	6	AF072506	AF072506 Homo sapi
45	532	83.8	157711	9	AC055738	AC055738 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX001024 635 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 6 from Patent WO902666.
ACCESSION AX001024
VERSION AX001024.1 GI:7241262

SOURCE
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 635)
AUTHORS Ott,C. and Bedin,F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 6 21 -JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
1..635 Location/Qualifiers
BASE COUNT 188 a 170 c 136 g 141 t
ORIGIN

Query Match 100.0%; Score 635; DB 6; Length 635;
 Best Local Similarity 100.0%; Pred. No. 3.3e-192;
 Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccgtatcttcaaacctcctgttaagttgtctctccagaatacaaacgttaaaacta 60
 |||||||
 Db 1 CCTGTATCTTTAAACCTCCTTTGTTAGTTGTCTCTCCAGATCAAAACTGTAAACTA 60
 OY 61 caaatgtcttccaatgagacacacagatgagatccatagactaaatccacgtgagacc 120
 |||||||
 Db 61 CAATATTGTTCTTCAATATGAGACACAGATGATCCATCACTAAGATCCACGTGGACCC 120
 OY 121 ctgagacggcctgtcagaccatgctccgaagttaaatgacatgaaagcaccctccgag 180
 |||||||
 Db 121 CTGGACCGGCGCTGTGAGCCATGCTCCGATGTTAATGACATTGAAGACACCCCTCCGAG 180
 OY 181 gaattcacaatgcacacacccctactatgcccgaatcagcggaagcaagtagagcgt 240
 |||||||
 Db 181 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGACCGGT 240
 OY 241 catcagcaaacctcccccaacagacacttgggtttctcgttgaaggggagactgaagagc 300
 |||||||
 Db 241 CATCAGCCAACTCCCCAACAGACACTTGGGTTTCTGTGAGAGGGGGACTGAGAGAC 300
 OY 301 aggaactagctggaattcctcagagcaagaatccctaaagctagcttggaagtgact 360
 |||||||
 Db 301 AGGACTAGCTGATTTCTTAGGCAACGAAGATCCCTTAAGCTAGCTGGAAAGTGACT 360
 OY 361 gcatcacccttaaaacatgaggtctgcaactagctcaacccgacaaatcagaagctc 420
 |||||||
 Db 361 GCATCCACCTCTAATACATGGGGCTTGCAACTAGCTACACCCGACCAATCAGAGAGCTC 420
 OY 421 actaaatgctaaatagagcaaaaatagaggtcaagaataagccaatcactatctgctg 480
 |||||||
 Db 421 ACTAAATGCTTAATTAGGCAAAAATAGAGGTAAAGAAATAGCCAATCTATTATTCCTG 480
 OY 481 agagcacaacggagaggaagaagatcggagataaaacccaagcatcgaacggcgaacgg 540
 |||||||
 Db 481 AGAGCACACGGGAGGAGACAGAGATCGGATATTAACCCAGGATTCGAGCGGCAACGG 540
 OY 541 caaccctcttggttccctcctctgtatgagcgctcgtcttcaactatctactct 600
 |||||||
 Db 541 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTCACCTATTTCACCTCT 600
 OY 601 actaaatctgcaactgaaaaaataaaaaa 635
 |||||||
 Db 601 ATTAAATCTTGCAACTGAAAAAATAAAAAA 635

RESULT 2
 AF127229 2004 bp mRNA linear VRL 11-AUG-1999

LOCUS AF127229 Multiple sclerosis associated retrovirus element clone Cl6
 DEFINITION pol-emb/3'LTR-like mRNA sequence.

ACCESSION AF127229
 VERSION AF127229.1 GI:5726294

KEYWORDS multiple sclerosis associated retrovirus element.
 SOURCE Multiple sclerosis associated retrovirus element.
 ORGANISM 1 (bases 1 to 2004)
 Viruses; Retroid viruses; Retroviridae.

REFERENCE 1 (bases 1 to 2004)
 Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,

Mallet,F., Mandrand,B. and Perron,H.
 Molecular cloning and characterization of MSRV-related sequences

associated with retrovirus-like particles
 Virology 260 (1), 1-9 (1999)

JOURNAL MEDLINE 99335590
 PUBMED 10405350

REFERENCE 2 (bases 1 to 2004)
 Komurian-Pradel,F., Paranhos-Bacala,G., Bedin,F.,

REFERENCE 2 (bases 1 to 2004)
 Komurian-Pradel,F., Paranhos-Bacala,G., Bedin,F.,

Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
 Mallet,F., Mandrand,B. and Perron,H.
 Direct Submission
 Submitted (10-FEB-1999) UMR103 CNRS, bioMerieux, 46, allée
 D'Italie, Lyon 69007, France

FEATURES
 source location/Qualifiers
 1..2004
 /organism="multiple sclerosis associated retrovirus
 element"
 /db_xref="taxon:89382"
 /clone="CL6"

misc_feature 1..2004
 /note="similar to pol-emb and 3'LTR region"

BASE COUNT 641 a 476 c 438 g 449 t

ORIGIN

Query Match 100.0%; Score 635; DB 14; Length 2004;
 Best Local Similarity 100.0%; Pred. No. 4.2e-192;
 Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cccgtatcttcaaacctcctgttaagttgtctctccagaatacaaacgttaaaacta 60
 |||||||
 Db 1370 CCTGTATCTTTAAACCTCCTTTGTTAGTTGTCTCTCCAGATCAAAACTGTAAACTA 1429
 OY 61 caaatgtcttccaatgagacacacagatgagatccatagactaaatccacgtgagacc 120
 |||||||
 Db 1430 CAATATTGTTCTTCAATATGAGACACAGATGATCCATCACTAAGATCCACGTGGACCC 1489
 OY 121 ctgagacggcctgtcagaccatgctccgaagttaaatgacatgaaagcaccctccgag 180
 |||||||
 Db 1490 CTGGACCGGCGCTGTGAGCCATGCTCCGATGTTAATGACATTGAAGACACCCCTCCGAG 1549
 OY 181 gaattcacaatgcacacacccctactatgcccgaatcagcggaagcaagtagagcgt 240
 |||||||
 Db 1550 GAAATCTCAACTGCACAAACCCCTACTATGCCCAATTCAGCGGGAAGCAGTTAGACCGGT 1609
 OY 241 catcagcaaacctcccccaacagacacttgggtttctcgttgaaggggagactgaagagc 300
 |||||||
 Db 1610 CATCAGCCAACTCCCCAACAGACACTTGGGTTTCTGTGAGAGGGGGACTGAGAGAC 1669
 OY 301 aggaactagctggaattcctcagagcaagaatccctaaagctagcttggaagtgact 360
 |||||||
 Db 1670 AGGACTAGCTGATTTCTTAGGCAACGAAGATCCCTTAAGCTAGCTGGAAAGTGACT 1729
 OY 361 gcatcacccttaaaacatgaggtctgcaactagctcaacccgacaaatcagaagctc 420
 |||||||
 Db 1730 GCATCCACCTCTAATACATGGGGCTTGCAACTAGCTACACCCGACCAATCAGAGAGCTC 1789
 OY 421 actaaatgctaaatagagcaaaaatagaggtcaagaataagccaatcactatctgctg 480
 |||||||
 Db 1790 ACTAAATGCTTAATTAGGCAAAAATAGAGGTAAAGAAATAGCCAATCTATTATTCCTG 1849
 OY 481 agagcacaacggagaggaagaagatcggagataaaacccaagcatcgaacggcgaacgg 540
 |||||||
 Db 1850 AGAGCACACGGGAGGAGACAGAGATCGGATATTAACCCAGGATTCGAGCGGCAACGG 1909
 OY 541 caaccctcttggttccctcctctgtatgagcgctcgtcttcaactatctactct 600
 |||||||
 Db 1910 CAACCCCTTTGGGTCCCTCCCTTTGTATGGCGCTCTGTTCACCTATTTCACCTCT 1969
 OY 601 actaaatctgcaactgaaaaaataaaaaa 635
 |||||||
 Db 1970 ATTAAATCTTGCAACTGAAAAAATAAAAAA 2004

RESULT 3
 AC019346 176095 bp DNA linear HTG 25-OCT-2001

LOCUS AC019346 Homo sapiens chromosome 18 clone RP11-497M7 map 18, WORKING DRAFT

DEFINITION SEQUENCE, 1 ordered pieces.

ACCESSION AC019346
 VERSION AC019346.4 GI:16418201

KEYWORDS HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 18, clone RP11-497M7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176095)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deaellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisenl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (02-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 25, 2001 this sequence version replaced g1:7230183. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L4296
Center clone name: 497_M_7

Summary Statistics
Sequencing vector: M13; M7815; 32% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 176087 bases at least Q40
Consensus quality: 176095 bases at least Q30
Insert size: 178000; agarose-fp
Insert size: 176095; sum-of-ctrls
Quality coverage: 11.5 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20.

NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 176095: contig of 176095 bp in length.
assembly-fragment.

Location/Qualifiers
1. 176095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-497M7"
/clone_lib="RPC1-11 Human Male BAC"

BASE COUNT 56205 a 30885 c 31099 g 57906 t
ORIGIN

Query Match 99.5%; Score 631.8; DB 2; Length 176095;
Best Local Similarity 99.7%; Pred. No. 1.1e-190;
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 cccgtatctttaaccctccgttgatgtgtctctccgaatacaactgttaacta 60
Db 150524 CCCGTATCTTTAAACCCCTGTTGAAGTTGTCTCTCCAGAAACCAACCGTAAACCTA 150465

61 caaatgtcttccaatgagacccaagatgagtcacatgactaagatccacgtgaccc 120
Db 150464 CAAATGTTCTTCAATATGAGACACACAGATGAGTCCATGACTAAGATCCACCGTGGACCC 150405

121 ctgagaccgacctgtagccacatgctcgaatgttaatgacatggaagcaccctccgag 180
Db 150404 CTGAGACCGCGCTGCTACCCATGCTCCGAGTGAATGACATTGAAAGCACCCCTCCGAG 150345

181 gaaatcctaactgacacacccctactatgcccgaatccagcggaagcaagttagagcgt 240
Db 150344 GAAATCTCAACTGCACACCCCTACTATGCCCCCAATTACACCGGACACGATTAGACGGGT 150285

241 catcagcacaacctcccaacagacattgggtttctcgttgaagaggggagctagagac 300
Db 150284 CATCAGCCAACTCCCAACAGACACTTGGGTTTTCCTGTTGAGAGGGGGGACCTAGAGAC 150225

301 aggaactagctggaattcctatagccaagaagaatccctaagcctagctggaagtgact 360
Db 150224 AGGACTAGCTGAGATTCTCTAGGCCAACGAAGAAATCCTTAAGCTGGGAAGGTGACT 150165

361 gcatcaccctctaactatgagcttgcaacttgctacacccgacccaatcagaagactc 420
Db 150164 GCATCCACCTCTTAAACATGGGGCTTGCACACTTACTACACCCGACCAATCAGAGACTC 150105

421 actaaatgctaactatgagcaaaaatagaggttaagaagaatagccaatcatctatgctg 480
Db 150104 ACTAAATGCTAATATTAGCAAAATATAGAGTAAAGAAATAGCCAAATCATTATGGCTG 150045

481 agagcacagcgaggaggaacaaagatcgagatataaaccaggaactcgagcggaagc 540
Db 150044 AGAGCACAGCGGAGGAGCAAGATCGGATATTAACCCAGGCAATGAGCGGCAACGG 149985

541 caaccccttggtgtccctcccttgatgagggcctgtttcatctatctatcctc 600
Db 149984 CAACCCCTTGGTGTCCCTCCCTTGTATGGGGCTGTGTTTCACCTATTTCACTCT 149925

601 attaatctgcaactgaaaaaataaaaaa 635
Db 149924 ATTAATCTTGCAACTGAAAAAATAAAAAAGAAAA 149890

RESULT 4
AC022171 210336 bp DNA 1linear HTG 26-MAR-2001
LOCUS Homo sapiens chromosome 18 clone RP11-407618, WORKING DRAFT
DEFINITION
SEQUENCE, 2 unordered pieces.
ACCESSION AC022171 GI:13270574
VERSION AC022171.18 GI:13270574
KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A., Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deaellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisenl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210336)

AUTHORS

Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P., Glukhova, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A. J., Oefner, P., Palm, C. J., Ramirez, D., Wilhelm, J., Yu, S., and Davis, R. W.

TITLE

Direct Submission
Submitted (26-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT

On Mar 10, 2001 this sequence version replaced gi:13122770.

----- Genome Center

Center: Stanford DNA Sequencing and Technology Development

Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: hum-info@sequence.stanford.edu

----- Project Information

Center project name: 758

Center clone name: RP11-407C18

----- Summary Statistics

Sequencing Vector: M13mp18; X02513; 97% of reads

Sequencing Vector: plasmid; plasmid_accession: 3% of reads

Chemistry: Dye-terminator Big Dye; 88% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 208312 bases at least Q40

Consensus quality: 208877 bases at least Q30

Consensus quality: 209157 bases at least Q20

Insert size: 199762; agarose-fp

Insert size: 210236; sum-of-contigs

Quality coverage: 8.7x in Q20 bases; agarose-fp

Quality coverage: 8.3x in Q20 bases; sum-of-contigs.

* NOTE: This is a working draft sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 66468: contig of 66468 bp in length

* 66469 66568: gap of unknown length

* 66569 210336: contig of 143768 bp in length.

Location/Qualifiers

1..210336

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/clone="RP11-407C18"

/clone.lib="RRCI human BAC library 11"

1..66468

/note="assembly_name:Contig44

clone_end:77"

66569..210336

/note="assembly_name:Contig45

clone_end:SP6"

BASE COUNT 67675 a 37228 c 37515 g 67816 t 102 others

ORIGIN

Query Match 99.5%; Score 631.8; DB 2; Length 210336;

Best Local Similarity 99.7%; Pred. No. 1.2e-190;

Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccctgtaccttaacctctgttaagttctctctccagaatcaaaactgtaaaacta 60

DB 58176 CCCTGTATCTTAACCTCTGTTAAGTTGCTCTCCAGATCAAAACGTAACATA 58235

QY 61 caaatgtcttcaaatgagacacagatgagtcacatgactaagaatccaccgtgacc 120

DB 58236 CAAATGTTCTTCAAAATGAGACACAGATGAGTCCATGATGATCACCCTGGACCC 58295

QY 121 ctgagccggcctgtgacgcatgtctcgaatgtaatgacatgtaaggaacccctccag 180

DB 58236 CTGACCGCGCTGCTAGCCCATGCTCCGATGTAAATGACATTTGAAGGACCCCTCCGAG 58355

QY 181 gaaatcctaactgcacaaacccctactatgcccactcaagcgggagagatgaagcgt 240

DB 58356 GAAATCTCACTGACCAACACCCCTACTATGCCCCCAATTCAGGGGGAAGCAGTTAGACCGGT 58415

QY 241 catcagccaacctcccacacagcacttgggttttcccttgaagaggggagactgaagac 300

DB 58416 CATCAGCCACACCTCCCAACAGCAGCTTGGGTTTCTCTTTAGAGGGGGGAGCTGAGAAC 58475

QY 301 aggaactagctgatttctcttaagcaagaagaatcccaagccctgagcgggaaagtgc 360

DB 58476 AGGACTACTGATGATTTCTTAAGCCCAACAGATCTCTTAAGCTTACTGCTGGAGAGTGACT 58535

QY 351 gcatccaccttaacacatgaggctgtgcaactagctacacccgacacatgaagagctc 420

DB 58536 GCATCCACCTCTTAACATGAGGGCTTGCAACTTACTGACACCCGACCAATCAGAGAGCTC 58595

QY 421 actaaatgctaataatgagcaaaataggaaggtlaagaaatagccaatcatctatgtcgt 480

DB 58596 ACTAAATGCTAATATGACCAAAATAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTG 58655

QY 481 agagacagcgggagagcaagagatcgagatataaacccagcattgagccggaacgg 540

DB 58656 AGAGCAGCAGCGGAGGAGCAAGATCGGATTAATTAACCCAGCATTGCAAGCGGCAACGG 58715

QY 541 caaccctcttgggtccctccctctgtatgagcgctctgtttcactctatctactct 600

DB 58716 CAACCCCTTTGGGTCCTCCCTCTTGTATGAGGGCCTGTGTTTCACTATTTCACACT 58775

QY 601 attaaatcttcaactgtaaaaaaataaaaaa 635

DB 58776 ATTAATCTTGCAACTGAAAAAAGAAAAA 58810

RESULT 5

AC010778 148724 bp DNA linear HTG 30-MAR-2000

LOCUS AC010778/c Homo sapiens clone RP11-2N15, WORKING DRAFT SEQUENCE, 8 unordered

DEFINITION pieces.

AC010778

AC010778.3 GI:7341824

VERSION HTG: HTGS_PHASE1: HTGS_DRAFT.

KEYWORDS human.

SOURCE ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 148724)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-2N15

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 148724)

JOURNAL Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Becker, R., Boguski, L., Bouckgeater, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Galagan, J., Gargyala, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, R., Macdonald, P., Marquis, N.,

McMan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

COMMENT Direct Submission

Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 30, 2000 this sequence version replaced gi:6087973.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Em., EMBL, SW., SWISSPROT, Tr., TREMBL, Wp., WORMEP, Information on the WORMEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the library RPCI-6 constructed by restriction digest. RP6-198C4 is from the further details see <http://www.choil.org/bacpac/home.htm>

VECTOR: pPAC4

IMPORTANT: This sequence is not the entire insert of clone RP6-198C4. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

FEATURES

source 1. 8339
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q13.3-21.2"
/clone="RP6-198C4"
/clone.lib="RPCI-6"

repeat_region 1..5641 repeat: matches 368..5930 of consensus"
repeat_region 5635..5673 repeat: matches 7419..7457 of consensus"
repeat_region 5664..5955 repeat: matches 8244..8523 of consensus"
repeat_region 5956..6136 repeat: matches 1..239 of consensus"
repeat_region 6135..6280 repeat: matches 239..324 of consensus"
BASE COUNT 2649 a 1846 c 1726 g 2118 t
ORIGIN

Query Match 91.2%; Score 579.2; DB 9; Length 8339;
Best Local Similarity 94.8%; Pred. No. 4.4e-174;
Matches 599; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

1 cccctacattcaactctgttaagtgtctcttcagagatacaaacgttaacta 60
|||||
5664 CCCTGTATCTTAACTCTCTGTAAAGTTGCTCTCCAGAACTGTATAACTA 5723
61 caaatgtcttcaaatgagacacagatgagatcactgaacacgctgaccc 120
|||||
5724 CAAATGTTCTTCAATGAGACCCAGATGCATGATCAATACACCCGCGATCC 5783
121 ctgagacgctgtcagcaccatgctcgaatgtaacatgaagcaccctccgag 180
|||||
5784 CTGGACACGCGCTGAGCCCATGCTCCAGTGTAAATGATGAGGACACCCCTCCGAG 5843
181 gaaatctaactgagacacccctactatgcccacatcagcgggagagatgaagcggt 240
|||||
5844 GAAATCTAACTGACAAACCCCTACTATGCCCAATTAAGCAGGAGCACTTAAGACAGT 5903
241 catcagcaaacctcccaacagcacttggtttctctgtgagaggggagctgagagac 300
|||||
5904 CATGAGCAACTCCCAACAGCACTTGGCTTTCTCTGTAGAGGCGGTACTGAGACAGC 5963
301 aggaactagctgattctctagggcaagaatcccttaagcctagctgaggagtgact 360
|||||
5964 AGGACTACTGATTTCTAGGCAACTAAGAACTCCCTAAGCCTAGTGGAGAGGTGACC 6023
361 gcatcaccctctaaacatgagggtctgcaacttagctcaccccgacacatagagaagctc 420
|||||

Db 6034 ACACCCACCTTTAAACATGAGGCGCTTGAACCTAGCTCACACCAACCAATAGAGAGCTC 6083
Oy 421 actaaatgctaatataggcaaaaatagagagtaaaagaatagccaatcatctatgctg 480
|||||
Db 6084 ACTAAATGCTATTAATGACCAAAAACAGAGAGTAAGAAATAGCCAAATCATCTATTGCTG 6143
Oy 481 agagacacgaggagagcaagagatcgggataaaaccagagcattcgagccgcaagc 540
|||||
Db 6144 AGAGCAGAGCTGAGAGGACAAAGATCGGAGATTAACCCAGGCAATTCGCGCAGACAG 6203
Oy 541 caaccctcttggtgcccctccctctgtatagggcgctctgtttactatcttactct 600
Db 6204 CAACCCCTTGGGTGCGCCCTCTCTGTATGAGGAGCTGTTTCACTATTTACCTCT 6263
Oy 601 attaaacttgcaactgaaaaaataaaaaa 632
|||||
Db 6264 ATTAATCTTGCACTGAAAAAATAATGCAN 6295

RESULT 7
AC073626/c 136901 bp DNA linear PRI 09-MAY-2001
LOCUS Homo sapiens BAC clone RP11-95P9 from 7, complete sequence.
DEFINITION AC073626
AC073626 GI:12863221
VERSION AC073626.7
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 136901)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 136901)
Nguyen,C., Haakenson,B. and Bauer,H.
The sequence of Homo sapiens BAC clone RP11-95P9
unpublished
REFERENCE 3 (bases 1 to 136901)
Waterston,R.H.
Aurthors
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 136901)
Waterston,R.H.
Aurthors
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 136901)
Waterston,R.
Aurthors
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 16, 2001 this sequence version replaced g1:11597125.

COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gac>
Contact: sapient@watson.wustl.edu
Summary Statistics
Center project name: H_NH0095P09

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRF/CH7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dpcpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTD-2023N18, 200 bp overlap: the clone sequenced to the right is CTA-250D13, 200 bp overlap. Actual start of this clone is at base position 127066 of CTD-2023N18; actual end is at base position 26937 of CTA-250D13.

There are polymorphic base differences between the clone RP11-95P9 and stolen data from a redundant clone RP11-636N5. The region from base position 29435 to 29729 consists of an insertion fragment from RP11-636N5, which was originally found deleted RP11-95P9.

FEATURES

source

1. 136901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-95P9"
/clone_lib="RPCT-11"
545..1157
/rpt_family="L1"
1195..1360
/rpt_family="L1"
1530..1791
/rpt_family="L1"
2763..33220
/note="similar to EST AA564655 (NID:g2336294) n125902.s1"
2972..3112
/rpt_family="MERL_type"
3118..3308
/rpt_family="MERL_type"
6682..6685
/rpt_family="MIR"
7038..7368
/rpt_family="MaLR"
7433..7802
/rpt_family="MaLR"
7811..7938
/rpt_family="(TA)n"
7945..7994
/rpt_family="(TA)n"
repeat_region
/rpt_family="AT_rich"
8005..8064
/rpt_family="(TATATC)n"
repeat_region
8067..8206
/rpt_family="MaLR"
9047..9246
repeat_region
/rpt_family="MaLR"
9227..9306
misc_feature

/note="similar to EST AU035546 (NID:g3718549)"
9289..9399
/rpt_family="MaLR"
9412..9494
/rpt_family="MaLR"
9495..9854
/rpt_family="MaLR"
9855..9928
/rpt_family="MaLR"
repeat_region
/rpt_family="Alu"
10530..10814
10815..10909
/rpt_family="(CATA)n"
12990..13351
/note="similar to EST AV755789 (NID:g10913637)"
12990..13414
/note="similar to EST AW955488 (NID:g8145171)"
13025..13434
/note="similar to EST AV756558 (NID:g10914406)"
13026..13433
/note="similar to EST AV710618 (NID:g10729924)"
13026..13414
/note="similar to EST BE888689 (NID:g10345245)"
13026..13425
/note="similar to EST AW975063 (NID:g8166267)"
13026..13433
/note="similar to EST AV717668 (NID:g10814820)"
13026..13433
/note="similar to EST AV762499 (NID:g10920347)"
13026..13433
/note="similar to EST AV701025 (NID:g10717355)"
13026..13222
/note="similar to EST BF348560 (NID:g11296155)"
13026..13411
/note="similar to EST AV755769 (NID:g10913617)"
13026..13433
/note="similar to EST AV760867 (NID:g10918715)"
13026..13405
/note="similar to EST BE277509 (NID:g9152479)"
13026..13408
/note="similar to EST AA532593 (NID:g2276847) n154e08.s1"
13026..13357
/note="similar to EST BE907116 (NID:g10400459)"
13028..13430
/note="similar to EST BF031653 (NID:g10739345)"
13031..13434
/note="similar to EST A1879643 (NID:g5553692) au56e10.x1"
13031..13433
/note="similar to EST AW157724 (NID:g6229125) au86a05.x1"
13032..13224
/note="similar to EST AA608523 (NID:g2456951) ae53b04.s1"
13035..13171
/note="similar to EST A1476367 (NID:g4329412) cm21d12.x1"
13038..13411
/note="similar to EST BF237771 (NID:g11151690)"
13040..13404
/note="similar to EST BE399397 (NID:g9334762)"
13053..13434
/note="similar to EST A1879261 (NID:g5553310) au56e10.y1"
13119..13435
/note="similar to EST BE710307 (NID:g10098572)"
13182..13413
/note="similar to EST BF034781 (NID:g10742493)"
14398..14434
/rpt_family="(TAA)n"
14453..14659
/rpt_family="MIR"
15131..15248
/rpt_family="(MER94)"
15255..15456
/rpt_family="MaLR"
15457..15494
/rpt_family="AT_rich"


```

OY 421 actaaatgtaattagcaaaaataaggagtaagaataagccaatcattctgctg 480
Db 83851 ACTAAATGCTAATTAGCAAAAACAGGAGTAAGAAATAGCAATCATCTATTGCTCT 83910
OY 481 agagccagagggagagacagatcggtatataaccaggcattcgagccgcaacg 540
Db 83911 AGAGCAGACAGAGAGGAGCAAGATCAGATATTAACCCAGGCTTCGAGCGGCAACAG 83970
OY 541 caaccctcttggtccccccttgatggcgctctgttctactatctactct 600
Db 83971 CAACCCCTTTGGGTCCCTCCCTTTATGGAGCTCTGTTTTCATCTATTCTACTCT 84030
OY 601 attaaatctgcactgaaaaaataaaaaa 635
Db 84031 ATTAATCTGCAACTCAAAAAAATAAAAAA 84065

RESULT 9
AL451130/c 212096 bp DNA linear HTG 08-FEB-2002
OCUS Homo sapiens chromosome 9 clone RP11-10G21, *** SEQUENCING IN
DEFINITION PROGRESS *** 9 unordered pieces.
ACCESSION AL451130 AC011776
VERSION AL451130.7 GI:18643763
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Clark, G.
AUTHORS Direct Submission
TITLE Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridge, UK. E-mail enquiries: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:18151469.
Draft Sequence produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: DA10621
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 59% of reads
Sequencing vector: plasmid; L08752; 40% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 206856 bases at least Q40
Consensus quality: 208786 bases at least Q30
Consensus quality: 210158 bases at least Q20
Insert size: 211296; sum-of-ctnigs
Insert size: 183186; 9.1% error; agarose-fp
Quality coverage: 9.78x in Q20 bases; sum-of-ctnigs
Quality coverage: 12.66x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2459: contig of 2459 bp in length
* 2460 2559: gap of 100 bp
* 2560 4643: contig of 2084 bp in length
* 4644 4743: gap of 100 bp
* 4744 6823: contig of 2080 bp in length

```

```

* 6824 9223: gap of 100 bp
* 6924 9152: contig of 2229 bp in length
* 9153 9252: gap of 100 bp
* 9253 198727: contig of 189475 bp in length
* 198728 198827: gap of 100 bp
* 198828 202432: contig of 3605 bp in length
* 202433 202532: gap of 100 bp
* 202533 206117: contig of 3585 bp in length
* 206118 206217: gap of 100 bp
* 206218 209532: contig of 3315 bp in length
* 209533 209632: gap of 100 bp
* 209633 212096: contig of 2464 bp in length.
FEATURES
source
1. .212096
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-10G21"
/clone_11b="RPC1-11.1"
1. .2459
/misc_feature
/note="assembly_fragment:00156"
2560. .4643
/note="assembly_fragment:00601"
4744. .6823
/note="assembly_fragment:01422"
6924. .9152
/note="assembly_fragment:01529"
9253. .198727
/note="assembly_fragment:02535"
198828. .202432
/note="assembly_fragment:03097"
202533. .206117
/note="assembly_fragment:03385"
206218. .209532
/note="assembly_fragment:04246"
209633. .212096
/note="assembly_fragment:04283"
BASE COUNT 63105 a 38564 c 41164 g 68462 t 801 others
ORIGIN
Query Match 88.9%; Score 564.6; DB 2; Length 212096;
Best Local Similarity 94.2%; Pred. No. 4.2e-169;
Matches 598; Conservative 0; Mismatches 34; Indels 3; Gaps 1;
OY 1 cccgtatctttaaaccctctctgttaagttgtctcttccaaatcaaacgttaaacta 60
Db 113775 CCGGTATCTTTAACTCCTCTGTTAAGTGTCTCTCCAGAAATCGAAGCTGTAA--TA 113719
OY 61 caattgtcttcaaatgagacacagatgagatgcatgactaacacgctggagacc 120
Db 113718 TATATGTTCTTCAAAATGAGACACAGATGCAATGATTAAGATTAACCGGAGACC 113659
OY 121 ctgagccgacctgctagccatgctcgcagtgtaatgacatggaagcaccctcccgag 180
Db 113658 CTGAGACGGCGCTGTAAGCCATGCTCGATGTAAACGACATTAAGGACACCCCTCTAG 113599
OY 181 gaaatctcaactgcaaacccctactatgcccacatctaacggggaagatgtagcggt 240
Db 113598 GAAATCTCACTGACCAACCAACTACTATGCCCAATTCAGCAGAGACGATTAGCGCGT 113539
OY 241 catcagcaactccccaagacacttggtcttctctgtgagagggagctgagagac 300
Db 113538 TGTACGCAACCTCCCAACAGCACTGGGTTTCTTTAGAGAGGGGAGCTAGAGAC 113479
OY 301 agagctagctgattctctagcaacgaaagaaatcccttaagccttagctggaagtgact 360
Db 113478 AGGACTAGCTGGAATTTCTAGGCGGAATTAAGAAATCCCTTAAGCTGAGGAGGACT 113419
OY 361 gcatccactttaaacaaggggcttgcaacttagcttaacacccgagcaataagagagctc 420
Db 113418 GCATCCACCTTTAAACACAGGGGCTTGCAACTTAGCTCACACCCCAACATCAGAGAGCTC 113359

```


QY	421	ACTAATAAGCCTAATTAAGGCAAAATAGGAGGTAAGGAAATAGCCATATCATTCATGCGCG	480
Db	113358	ACTAATAAGCCTAATTAAGGCAAAATAGGAGGTAAGGAAATAGCCATATCATTCATGCGCG	113299
QY	481	AGAGCAACGCGGAGGAGGCAAGAGATCGGATATAAACCAGGATTCAGCCGCAACG	540
Db	113298	AGAGCAACGCGGAGGAGGCAAGAGATTCAGGATATAAACCAGGATTCAGCCGCAACG	113239
QY	541	CAACCCCTTTGGTGGCTCCCTCCTTGTATGAGGAGCTGTGTTTCATTCATTCACCTCT	600
Db	113238	CAACCCCTTTGGTGGCTCCCTCCTTGTATGAGGAGCTGTGTTTCATTCATTCACCTCT	113179
QY	601	ATTAATCTCTGCAACTGCACTGCAAAAAAATTTTTTTTTTTT	635
Db	113178	ATTAATCTCTGCAACTGCACTGCAAAAAAATTTTTTTTTTTT	113144
RESULT	10		
AC072023/c			
LOCUS			
DEFINITION	AC072023	175040 bp	DNA
			linear
VERSION	AC072023.4		HTG 31-JAN-2002
KEYWORDS	AC072023.4	GI:17432828	WORKING DRAFT SEQUENCE, 7
SOURCE	HTG: HTGS_PHASE1; HTGS_DRAFT.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 175040)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oshan,F.R., Allen,C., Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonito,D., Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Butney,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., David,M.L., Davis,C., Davy-carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Degen,A.L., Ding,Y., Dinh,K.J., Douthwaite,K.J., Draper,H., Dumain-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gottell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudh,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kueshl,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Meshawari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Messey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K., Morgan,M., Morris,S., Moser,M., Neel,D., Newtson,T., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Ogun,M., Okunolu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primeau,E., Pu,L.L., Qules,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshantali,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sulton,A., Svaitek,A., Taber,P., Tameris,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanli,K., Vasquez,L., Vera,V., Villalona,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.		
TITLE	Unpublished		
JOURNAL	Direct Submission		
REFERENCE	2 (bases 1 to 175040)		

```

AUTHORS
TITLE
JOURNAL
COMMENT

Morley, K.C.
Direct Submission
Submitted (07-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 10, 2001 this sequence version replaced gi:9438656.

-----
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: HBEC
Center clone name: Rp11-3j2
-----
Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList
Consensus quality: 177002 bases at least Q40
Consensus quality: 181942 bases at least Q30
Consensus quality: 185128 bases at least Q20
Estimated insert size: 180460; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ip estimation
Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 47110: contig of 47110 bp in length
* 47111 47210: gap of unknown length
* 47211 80061: contig of 32851 bp in length
* 80062 80161: gap of unknown length
* 80162 112177: contig of 32016 bp in length
* 112178 112277: gap of unknown length
* 112278 140321: contig of 28044 bp in length
* 140322 140421: gap of unknown length
* 140422 157647: contig of 17226 bp in length
* 157648 157747: gap of unknown length
* 157748 168444: contig of 10697 bp in length
* 168445 168544: gap of unknown length
* 168545 175040: contig of 6496 bp in length.
*
Location/Qualifiers
1. 175040
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="Rp11-3j2"

BASE COUNT 53880 a 32737 c 32187 g 55620 t 616 others

ORIGIN

Query Match 88.2%; Score 560.2; DB 2; Length 175040;
Best Local Similarity 93.8%; Pred. No. 1e-167;
Matches 594; Conservative 0; Mismatches 38; Indels 1; Gaps 1

OY 1 cccgtgataccttaacccctcgttaagtttgctcttccagaatacaaaactytaaacta 60
|||||
Db 123368 CCCGTGATCTTTAACCTCTCTGTGAAGTTGTCCTTCACAGAAATCAAAATGTTAAACATA 123309

OY 61 caaatgtcttcaaatgagacacagatgagtcatacagacaaatcacacgtygagacc 120
|||||
Db 123308 CAAATGGGTCTTCAAAATGAGACCCCAAGATGCGTCATGATCAAGATCCACACAGACC 123249

OY 121 ctgagaccgctctgtagccatgctccagatgttaatgacatgaaagagacccctccgag 180
|||||
Db 123248 CTGAGACGTGCTCTAGCCCATGATCCGATGTTAATGACATCAAAAGCACCCCTCCCAAG 123189

```


QY 181 gaatctcaactgcacacccctactatgccccaattcagcgggaagcagttagacgct 240
|||||
Db 123188 GAATCTCAACTGCGCCACACCCCTACTATGCCCAATTCAGCAGAACACAGTACAGAGT 123129
QY 241 catcagccaacccctcccaacagcacttggcttctcgttgaagaagggaagcagac 300
|||
Db 123128 CATTGGCCCAACCCGCCAACAGCAGCTTGATCTCTGTTGAGAGGGGGGACTGAGAGAT 123069
QY 301 aggaactagcgggatttcttggcccaagaagaatccctaagccctagcttgggaagttact 360
|||||
Db 123068 AGGACTAGCTGGATTCTCTAGGCCAATCAATCCCTAAGCTTACTGCGAAGGTACCC 123009
QY 361 gcatccacccttaacataagggtgcttgaacttaagctcacaccgacccaatcaagaagctc 420
|||||
Db 123008 GCATCCACCTTTAAACACGGGGGCTCGCACTAGCTCAGCCACCAATCAGAGACTC 122949
QY 421 actaaatgtcatttagcgaanaaataaggaggtaaagaataagccaatcatctatgtcctg 480
|||||
Db 122948 ACTAAATGCTAATTAGGCAAAAACAGAGGTAAA-TAAATAGCCAAATCATATTGCTG 122890
QY 481 agagcacagggagggagacaagatcgggatataaccggagcattgaagccggcaagc 540
|||||
Db 122889 AGACACAGCGGGAGGAGACAGGAGATATAAACAGGCAATGCGAGCTGGCAATGG 122830
QY 541 caaccccttgggtccctccctctgtatggcgctctgtttcactctatctactct 600
|||||
Db 122829 CAACCCCTTTGGGTCCCTTCCTGTATGGGAGCTCTGTTTTCACCTCATTTACTCT 122770
QY 601 attaaacttgcacactgaaaaaagaaaaa 633
|||||
Db 122769 ATTAATCTTGCAACTGCAAAAAA 122737

RESULT 11
AP001538/c 174019 bp DNA linear PRI 25-MAR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
LL56-APP region, complete sequence.
ACCESSION AP001538
VERSION AP001538.1 GI:7328982
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B853K11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174019)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
Published only in Database (2000) In press
2 (bases 1 to 174019)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (23-MAR-2000) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.c.riken.go.jp) URL:http://ngp.gsc.riken.go.jp/
Tel:81-42-778-9923, Fax:81-42-778-9924)
Location/Qualifiers
1. 174019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B853K11"
/map="21q21.1-q21.2"
BASE COUNT 53303 a 31416 c 32248 g 57052 t
ORIGIN

Query Match 87.7%; Score 556.6; DB 9; Length 174019;
Best Local Similarity 93.7%; Pred. No. 1,Se-166;

Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 cccggtatctttaaaccctccctgttaagtgttctctccagaatccaactgttaacta 60
|||||
Db 36710 CCCGTATCTTTAAACCTCCCTGTTAACTTGTCTTCCAGAAACAAACGTAAACTA 36651
QY 61 caatttcttcaaatggagaccagatgagatcagatcagatcagatcagatcagatcagatc 120
|||||
Db 36650 CAATTAAGTCTTCAATAGGGGCCCCAGATGATGATGATGATGATGATGATGATGATG 36591
QY 121 ctggacggcctctgtagcccatgtctcgatgttaatgacattgaaggacccctccag 180
|||||
Db 36590 CTAGACGTGCGCTGTAGCCCATGCTCGATATTAATGACATGAAAGCACCCCTCCCAAG 36531
QY 181 gaatctcaactgcacacccctactatgtcccaatcagcgggaagcagttagacgct 240
|||||
Db 36530 GAAATCAACACTGCCCAACCCCTAATGCCCCCAATTCAGAGGAGAACAGTAAAGGTGT 36471
QY 241 catcagccaacccctcccaacagcacttgggttctcctgttgaagaagggaagcagac 300
|||||
Db 36470 CATTGGCCCAACCTCCCAACAGCAGCTTGGGTTTCTGTTGAGGGGGGCTAGAGAC 36411
QY 301 aggaactagctggatttctcctagggccaagaatccctaagcctagctgggaagttact 360
|||||
Db 36410 AGGACTAGCTGGATTCTCTAGGCTGACTAAGAAATCCCTAAGCTTGGGAAAGTGACC 36351
QY 361 gcatccacccttaacataagggtgcttgaacttaagctcacaccgacccaatcaagaagctc 420
|||||
Db 36350 GCATCCACCTTTTAAACATGGGGGCTCGCAACTTACTCTACACCCAGTCAGTAGAGACTC 36291
QY 421 actaaatgtcatttagcgaanaaataaggaggtaaagaataagccaatcatctatgtcctg 480
|||||
Db 36290 ACTAAATGCTAATTAGGCAAAAACAGAGGTAAAGAAATGCCAATCATATTATGCTGCT 36231
QY 481 agagcacagcgggagggacaagatcgggatataaaccagcatctgagccggcaagc 540
|||||
Db 36230 AGACACAGCGAGGAGGAGACAGGATGATGATGATGATGATGATGATGATGATGATGATG 36171
QY 541 caaccccttgggtccctccctctgtatggcgctctgtttcactctatctactct 600
|||||
Db 36170 CAACCCCTTTGGGTCCCTTCCTGTATGGGAGCTCTGTTTTCACCTCATTTACTCT 36111
QY 601 attaaacttgcacactgaaaaaagaaaaa 633
|||||
Db 36110 ACTAATCTTGCAACTGCA 36092

RESULT 12
AP001674/c 340000 bp DNA linear PRI 30-MAY-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 18/105.
ACCESSION AP001674 AL163219 BA000005
VERSION AP001674.1 GI:7768666
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Seede,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Seede,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.

TITLE
 JOURNAL
 COMMENT
 On May 30, 2000 this sequence version replaced g1.7717271.
 The chromosome 21 mapping and sequencing consortium consisting of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagamihara 228-8555, Japan,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *
 Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Molecular Biology, * Tokyo
 160-8582, Japan,
 * e.mail: nshimizu@med.keio.ac.jp
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
 info.genomegbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Innestrasse 73, D-14195 Berlin, Germany,
 * e.mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.tz-berlin.mpg.de/
 AL163219; Submitted (10-Apr-2000).
 Location/Qualifiers
 1. 340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 <1. 6149
 /note="Accession No. AP001537"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q21.1"
 /clone="R749F18, 5' partial"
 /clone_lib="RPCT-11 BAC library"
 <1. 165231
 /note="Accession No. AP001538"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q21.1"
 /clone="R853K11, 5' partial"
 /clone_lib="RPCT-11 BAC library"
 15. 444
 /note="LIM44A"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 555. 3531
 /note="LIM44A"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED

repeat_region	3770. .3876 /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED complement(5645. .6178) /note="LlM2" /rpt_family="LINE/L1" /rpt_type=DISPERSED 6181. .6534 /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 8021. .8370 /note="THE1B" /rpt_family="LTR/MaLR" /rpt_type=DISPERSED 15810. .15851 /note="(CA)n" /rpt_family="Simple_repeat" /rpt_type= tandem complement(16628. .16697) /note="MER53" /rpt_family="DNA" /rpt_type=DISPERSED 16638. .16715 /note="MER53" /rpt_family="DNA" /rpt_type=DISPERSED 18847. .19585 /note="L2" /rpt_family="LINE/L2" /rpt_type=DISPERSED 19795. .19917 /note="(TA)n" /rpt_family="Simple_repeat" /rpt_type= tandem 20023. .20166 /note="(TA)n" /rpt_family="Simple_repeat" /rpt_type= tandem 20289. .20448 /note="(TA)n" /rpt_family="Simple_repeat" /rpt_type= tandem 20467. .20542 /note="(TA)n" /rpt_family="Simple_repeat" /rpt_type= tandem complement(22160. .22028) /note="LlM3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(22142. .22448) /note="Alu5x" /rpt_family="SINE/Alu" /rpt_type=DISPERSED complement(22519. .22891) /note="MER32A" /rpt_family="LTR/MER4-group /rpt_type=DISPERSED 22976. .23339 /note="LlM3" /rpt_family="LINE/L1" /rpt_type=DISPERSED complement(23344. .23519)
---------------	--

```

/Note="MER4C"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
repeat_region complement(23644..23861)
/Note="MER4C"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
repeat_region complement(23925..25091)
/Note="MER52A"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
repeat_region complement(25227..25672)
/Note="MER4A2"
/rpt_family="LTR/MER4-group"
/rpt_type=DISPERSED
repeat_region complement(25682..25987)
/Note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(26744..26849)
/Note="MER46B"
/rpt_family="DNA/MER2_type"
/rpt_type=DISPERSED
repeat_region complement(26852..27629)
/Note="LTR17"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region complement(27630..27921)
/Note="HERV17"
/rpt_family="LTR/Retroviral"

Query Match      87.7%; Score 556.6; DB 9; Length 340000;
Best Local Similarity 93.7%; Pred. No. 1.7e-166;
Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 cccctgatacttaacctctctgttaagttgtctcttcgaatcaaaactgtaaacta 60
DB 27921 CCTGTATCTTTAACTCCCTGTTAGTTGTCTCTCCAGAAATCAAAACTGTAAACTA 27862
QY 61 caattgtcttcaaatgagacccagatgagtgatccatgactgaatccacgctgagccc 120
DB 27861 CAAATAGTTTCAATAGGGGGCCCCCAGATCAGTCCATGACTAGATCAACCGGGGCC 27802
QY 121 ctgagccgctctgtagcccatgctccgatatgataatgacattgaagcaccctccag 180
DB 27801 CTAGACAGGCGCTGCTAGCCCATGCTCGATATTAATGACATCGAAGCACCCTCCAG 27742
QY 181 gaaatctcaactgacacacccctactatgcccgaatcagcgaggaagcagtagagcg 240
DB 27741 GAAATCACAACCTCCCAACCCCTACTATGCCCAATTCACAGCAAGCACTAGAGCTGT 27682
QY 241 catcagccaaactcccaaaagacattggtttctcgtgtgagagggagacgaagac 300
DB 27681 CATTCGCAACCTCCCAACAGACATGAGTTTCTGTTGAGGGCGGTTACTGAGAGAC 27622
QY 301 aggaactagctgatttctctagccaagaagaatccctaagcctagctgaggaagtgact 360
DB 27621 AGACATGAGCGGATTCCTTAGGCTAGACTAGAAATCCCTAAGCTAGCTGGGAAGGTACC 27562
QY 361 gcatcaactctaaacattgggcttgtaactgtacacccgacccaatcagagagctc 420
DB 27561 GCATTCACCTTTAAACATGGGGCTCGCACTAGCTACACCAACCAAGTCAGAGAGCTC 27502
QY 421 actaaatgtctaataggcaaaaataggagaggtaaagaatagccaatcatctattgctg 480
DB 27501 ACTAAATGCTTAATAGGCAAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTG 27442
QY 481 agagcacaagcgagagacaaagatcgagatataaacccagagcatctgagcgcaagcg 540
DB 27441 AGAGCACAAGCAGGAGGAGCAAGAGTATGATTAACCCAGGCAATTCAACTGGCAACGG 27382
QY 541 caaccctcttggtccctccctctgtatgagcgctctgttttactattactctct 600

```

```

DB 27381 CAACCCCTTTGGGTCCTCCCTGCTGCTGTATGAGAGCTGTGTTTCACCTATTCACCTC 27322
QY 601 attaatctgcaactgaa 619
DB 27321 ACTAAATCTTGCACTGCA 27303

RESULT 13
AP001545
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-762624 map 18q21, WORKING
DRAFT SEQUENCE, 13 unordered pieces.
ACCESSION
AP001545
VERSION
AP001545.3 GI:9188503
KEYWORDS
HTG, HTGS, PHASE1, HTGS-DRAFT.
SOURCE
Homo sapiens DNA, clone:RP11-762624.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149755)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 149,755 genomic DNA of 18q21
Published Only in Database (2000) In press
2 (bases 1 to 149755)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (24-MAR-2000) Masahito Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Katsato, Sagami-hara, Kanagawa 228-8555, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8117383.

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: Humdrat18
Center clone name: RP11-762624
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Big-amerham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 14596 bases at least Q40
Consensus quality: 147102 bases at least Q30
Consensus quality: 147971 bases at least Q20
Insert size: 14855; sum-of-ctgifs
Quality coverage: 10.98x in Q20 bases; sum-of-ctgifs

NOTE: This is a 'working draft' sequence. It currently consists of
13 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 44828 contig of 44828 bp in length
44929 82602 contig of 37674 bp in length
82703 101344 contig of 18642 bp in length
101445 116205 contig of 14761 bp in length
116306 123784 contig of 7479 bp in length
123885 130355 contig of 6471 bp in length
130456 133619 contig of 3164 bp in length
133720 137416 contig of 3697 bp in length
137517 141540 contig of 4024 bp in length
141641 144351 contig of 2711 bp in length
144452 146667 contig of 2216 bp in length
146768 148219 contig of 1452 bp in length
148320 149755 contig of 1436 bp in length.

* NOTE: This is a 'working draft' sequence. It currently

```

* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 44828: contig of 44828 bp in length
* 44829 44928: gap of 100 bp
* 44929 82602: contig of 37674 bp in length
* 82603 82702: gap of 100 bp
* 82703 101344: contig of 18642 bp in length
* 101345 101444: gap of 100 bp
* 101445 116205: contig of 14761 bp in length
* 116206 116305: gap of 100 bp
* 116306 123784: contig of 7479 bp in length
* 123785 123884: gap of 100 bp
* 123885 130355: contig of 6471 bp in length
* 130356 130455: gap of 100 bp
* 130456 133619: contig of 3164 bp in length
* 133620 133719: gap of 100 bp
* 133720 137416: contig of 3697 bp in length
* 137417 137516: gap of 100 bp
* 137517 141540: contig of 4024 bp in length
* 141541 141640: gap of 100 bp
* 141641 144351: contig of 2711 bp in length
* 144352 144451: gap of 100 bp
* 144452 146667: contig of 2216 bp in length
* 146668 146767: gap of 100 bp
* 146768 148219: contig of 1452 bp in length
* 148220 148319: gap of 100 bp
* 148320 149755: contig of 1436 bp in length.

FEATURES

Location/Qualifiers

1..149755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-762624"
1..44828
misc_feature
/note="assembly_fragment"
44929..82602
misc_feature
/note="assembly_fragment"
82703..101344
misc_feature
/note="assembly_fragment"
101445..116205
misc_feature
/note="assembly_fragment"
116306..123784
misc_feature
/note="assembly_fragment"
123885..130355
misc_feature
/note="assembly_fragment"
130456..133619
misc_feature
/note="assembly_fragment clone_end:T7 vector_side:left"
133720..137416
misc_feature
/note="assembly_fragment"
137517..141540
misc_feature
/note="assembly_fragment clone_end:SP6 vector_side:right"
141641..144351
misc_feature
/note="assembly_fragment"
144452..146667
misc_feature
/note="assembly_fragment"
146768..148219
misc_feature
/note="assembly_fragment"
148320..149755
misc_feature
/note="assembly_fragment"
BASE COUNT 42913 a 31173 c 31787 g 42682 t 1200 others
ORIGIN

Query Match 86.9%; Score 551.8; DB 2; Length 149755;
Best Local Similarity 91.8%; Pred. No. 4.9e-165;
Matches 583; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 1 cccgtatccttaaacctcccttgatgaattgtctctctccagaataaactgtaaaccta 60
|||||
Db 67067 CCTCTATCTTTAACTCTCTTTAACTTTCTCTCCAGATCAAGCTGTAAACTTA 67126
Oy 61 caaatgtcttcaaatgagacacagatgagtgacatgactgaatgaatccacgtgacc 120
|||||
Db 67127 CAATCATCTTCTCAATGAGAGCCCAAGATGACATGATCAATGATTTACACAGACC 67186
Oy 121 ctgagccgctctctagccatctcgcagtgttaatgacatgagacccctccgag 180
|||||
Db 67187 CTGACACAGCGCTCGTAGCCATGCTCCGATGTATTGACATGAGAGCAACCTCCAA 67246
Oy 181 gaaatctcaactgcaaacctccatctatgcccccaatcaagcggaagcagttagaagcgt 240
|||||
Db 67247 GAAATCTCACTGACCAACCCCTACTATGCCCCCACTTACAGCAAGAGATTTAGGCGCT 67306
Oy 241 catagccaactccccaacagacacttggtttctctgttgaagaaggggagctagaagac 300
|||||
Db 67307 CATGACCAACCTCCCAACACGACTTGGGTTTCTGTCAGAGGGGCTACTGAGAGAC 67366
Oy 301 aggaactagctgattctcttaagcacaagaaagatccctcaagcctgctggaagtgact 360
|||||
Db 67367 AGGACTAGCTGGATTCTCTAGGCTGACTAAGATCCCTAAACCTAGCTGGAGAGTACT 67426
Oy 361 gcatccaccttaaacatgagggcttgcaactagctcaaccccgaccatcaagagctc 420
|||||
Db 67427 GCATTCACTTTAAACAGAGGCTTGCAACTTACTGACACTGACATCAGAGAGCTC 67486
Oy 421 actaaatgcttaattagcacaataatagaggttaagaataagcacaatcattgctgctg 480
|||||
Db 67487 ACTAAATGCTTAATTAGCCAAAACAGAGGTAAAGAAATAGCCAAATATCTATTGCTTG 67546
Oy 481 agagacagcgaggaagcaagagatcggagataaaaccagagcattcgagccggaacg 540
|||||
Db 67547 AGAGCACAGTGGAGGAGGACAGAGATCAGATATAAACCCAGGACTTACGCGCAAG 67606
Oy 541 caaccctcttggtgccctccctctgtatgagggcctctgttttaacctattcaacct 600
|||||
Db 67607 CAACCCCTTTGGGTCCTCCCTTTGTATGGAGCTGTGTTTCATCTTAATATCTT 67666
Oy 601 attaatcttgcaactgagaaaaaagaaaaa 635
|||||
Db 67667 TCAACTGCAAAAATTAATAATTAATAATTAATAA 67701

RESULT 14
AC090313 152775 bp DNA linear HTG 21-JUL-2001
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-762624 map 18, WORKING DRAFT
ACCESSION AC090313
VERSION AC090313.2 GI:14971265
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 152775)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-762624
JOURNML Unpublished
REFERENCES 2 (bases 1 to 152775)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camara,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,J., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., Laroque,K., Lamazares,R., Landers,T.,
Lehocky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldim,J., Meneus,L., Mihova,T., Mlenga,Y.,

COMMENT

On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.

FEATURES

SOURCE

1. 163803
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"
BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match

Best Local Similarity 86.9%; Score 551.8; DB 9; Length 163803;
Matches 577; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 cccctatcttaaacctcctctgttaagttgtctcttcagaatcaaaactgtaaacta 60
|||||
Db 117348 CCTGTATCTTTAACTTCTTGTAGTTTGTCTCTCCAGATTGAACTGTAAAGCTA 117289
QY 61 caaatgtctcaaatgagacacagatggagtcacagactaagatccaccgttgacc 120
|||||
Db 117288 CAAATAGTTCTTCAAAATGGAACCCAGATGAGTCATGACTAAATCTACCGTGACCC 117229
QY 121 ctggacggcctgtcagcccatgtccgatttaagacattgaagcaccctccgag 180
|||||
Db 117228 CTGGACCGCCTGTAGACTATGCTGTATTAATGACATTGAATGACCCCTCCGAG 117169
QY 181 gaaatcctcaactgcagaacccctactatgcccccaatlcagcgggaagcagttagaagcgt 240
|||||
Db 117168 GAAATCTCACTGCAACACCCCTACTACATCCAAATTCAGTAGAGCAAGTTAGACAGT 117109
QY 241 catcagccaacctcccaacacagcactgtgttctcgtltgagagggggaactgaagac 300
|||||
Db 117108 TGTGAGCCCAACCTCCCAACAGTACTGCTGTTCTGTTGAGAGGTGAGACTGAGAGAC 117049
QY 301 aggaactagcttgatttctcaggaacagaagaatccctaagcctagctgggaagtgact 360
|||||
Db 117048 AGGACTAGCTGGATTCTTAGCTGACTAAGATCCCTAAGCCTAGCTGGGAAGGTGACC 116989
QY 361 gcatccacctctaaacaatgggagctgtcaactagctcacaccggaaccaatcagaagactc 420
|||||
Db 116988 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGCAATCAGAGAGCTC 116929
QY 421 actaaatgtcaatcagcaaaaataggagtgaaagaatatgccaatcactatgtcctg 480
|||||
Db 116928 ACTAAATGTCAATCAGCAAAACAGAGGTAAGCAATAGCCAAATCATCTATTGCTG 116869
QY 481 agagacacggcgaggagcaagagatcgagatataaacccaagcattcagccggcaagcg 540
|||||
Db 116868 AGAGCACAGCGGAGAGCAAGAGATTGGATATTAACCCAGCATTCAGCCAGCAACAG 116809
QY 541 caaccctcttgggtccctcccttctgtatggcgctctgttctcaactatctcaactct 600
|||||
Db 116808 CAACCCCTTGGGTCCCTCCCATGTATGGAGCTCTGTTTTCACCTCTATTTCACCTCT 116749
QY 601 attaaatctgcaactgaa 619
|||||
Db 116748 ATTAAATCTGCAACTGCA 116730

Search completed: June 20, 2002, 09:43:49
Job time: 10754 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 08:18:17 ; Search time 3276.58 Seconds
(without alignments)
6100.565 Million cell updates/sec

Title: US-09-319-156a-9
Perfect score: 1481
Sequence: 1 atgcgcctccctatcatcac.....gtctacaatgaaccacca 1481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647	43.7	723	9	AUI38405 AUI38405
2	629.4	42.5	631	12	AQ261133 CTRBT-E1-
3	628.6	42.4	702	9	AUI38097 AUI38097
4	609.8	41.2	714	10	BE734284 BE734284
5	536.4	36.2	586	9	BE019603 BE019603
6	535.2	36.1	790	10	BG572445 BG572445
7	390.2	26.3	427	9	AA776439 AA776439
8	389.8	26.3	718	9	AL601847 AL601847
9	371.4	25.1	586	12	AQ776289 AQ776289
10	356.4	24.1	459	10	R68740 R68740
11	356.4	24.0	608	12	AQ725469 HS_5403_A
12	355.2	24.0	411	10	BG004254 RC6-GN007
13	349.6	23.6	382	10	B1052569 RC5-GN028
14	349.4	23.6	376	10	B1056595 RC5-GN028
15	347.4	23.5	373	10	B1052559 RC5-GN028
16	347.4	23.5	373	10	B1052567 RC5-GN028
17	346.8	23.4	374	10	B1056593 RC5-GN028

C 18	344.2	23.2	373	10	B1052561
C 19	344.2	23.2	373	10	B1052571
C 20	344.2	23.2	373	10	B1052572
C 21	344.2	23.2	373	10	B1056589
C 22	344.2	23.2	379	10	B1052566
C 23	342.6	23.1	373	10	B1052570
C 24	323.2	21.8	517	12	AQ170271 HS_3181_A
C 25	317.4	21.4	345	10	B1056592
C 26	309.4	20.9	346	10	B1053386
C 27	308.6	20.8	385	12	AQ062249 CTT-HSP-2
C 28	308.2	20.8	367	10	BF990707 RC6-GN007
C 29	304.8	20.6	379	10	D78692 HUM503H08B
C 30	289.6	19.6	334	10	B1053385
C 31	281.6	19.0	313	10	B1056596
C 32	274.8	18.6	322	10	BF990902 RC6-GN007
C 33	268.8	18.1	374	10	T93661 Y60590_r1
C 34	265.4	17.9	315	10	B1055459 RC5-GN028
C 35	265.4	17.9	315	10	B1056080 RC5-GN028
C 36	265.4	17.9	331	10	B1055458 RC5-GN028
C 37	265.4	17.9	331	10	B1056079 RC5-GN028
C 38	264.8	17.9	300	10	B1055456 RC5-GN028
C 39	264.8	17.9	300	10	B1056077 RC6-GN007
C 40	262	17.7	302	10	BG004247 RC6-GN007
C 41	258	17.4	317	10	B1056590 RC5-GN028
C 42	249.8	16.9	289	10	B1053219 PMO-GN021
C 43	249.6	16.9	43	10	B1053220 PMO-GN021
C 44	240.8	16.3	433	10	N77302 Y43603_r1
C 45	238.2	16.1	449	10	N58614 YV56h11.r1

ALIGNMENTS

RESULT 1
AUI38405
LOCUS AUI38405 723 bp mRNA linear EST 25-OCT-2000
DEFINITION AUI38405 PLACE1 Homo sapiens CDNA clone PLACE1008489 5', mRNA
sequence.

ACCESSION AUI38405
VERSION AUI38405.1 GI:10999926
KEYWORDS EST.

ORGANISM

SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 723)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.

TITLE

Unpublished (2000)
HRI human CDNA project
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952

COMMENT

Email: genomics@hri.co.jp
HRI human CDNA Project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; CDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source
1..723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1008489"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pMT185FL3"
BASE COUNT 210 a 181 c 130 g 197 t 5 others
ORIGIN

Db 601 AGCAAGAGTCTAGGACAGACTAGTACTGCG 631

RESULT 3
LOCUS AUI38097 702 bp mRNA linear EST 25-OCT-2000
DEFINITION AUI38097 PLACE1 Homo sapiens cDNA clone PLACE1007839 5', mRNA
sequence.
ACCESSION AUI38097
VERSION AUI38097.1 GI:10999618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1007839"
/clone_1id="PLACE1"
/issue_type="placenta"
/note="Vector: pME18SFL3"
BASE COUNT 206 a 173 c 131 g 189 t 3 others
ORIGIN

Query Match 42.4%; Score 628.6; DB 9; Length 702;
Best Local Similarity 94.6%; Pred. No. 1.5e-156;
Matches 660; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

Y 633 agaaataaacacacactccgttttagtagtagacactctgtttccaatctggaataaccac 692
Db 1 AGAATAAAGACCACTTCCGTTTAAAGTAGACCTCTTGTCCCAATCTGGAAATTAACCA 60

Y 693 taactcaaacctcacctgtgttaaaattagcaactatagaacacacacagctcccaatg 752
Db 61 TACCTCAAACTCACTCTGTGTAAATTTAGCAATACATACACAACTCCCAATG 120

Y 753 catagagtggttaacacactccacacagaaatagctgtcctaacctcagaatattttgt 812
Db 121 CATAGAGTGGTAACCTCTCCACACAAATAGTGTGCTTACCTCAGGAATATTTTGT 180

Y 813 ctgtgtgtaactcaagctatcatgtttgaatgagctctcctcagaatcatgctcctc 872
Db 181 CTGTGTGTAACCTCAAGCTATCTGTGTGAATGGCTCTTCAGAACTATGTGCTCTCTC 240

Y 873 attcttagtgccctctatagacacatctacacgaacaaattatatacatcatgtcc 932
Db 241 ATTCTTAGTGCCCTCTAAGACATCTACACGAACAAATTTATACAGTTATGTATATC 300

Y 933 taagccccaacaaagaagatccatctcctctctgttttctcaagcagagtgctag 992
Db 301 TAAGCCCGCAACAAAGATACCATCTCTCTTTTATAGAGAGAGAGGTCTAGG 360

Y 993 cagactagttactgtgcatgtgcagatcacacactctactcagttctactacaactatc 1052

Db 361 TGCAGTAGTACTGSCATTGGGGGTATCATCAACCTCTACTGAGTTCTACTACAACTATC 420

Y 1053 tcaagaataaatgtgtgacatggaacaggtcacgtacccctgtgtccacctgcaatca 1112
Db 421 TCAGAACTAAATGGGATGGAACGGGTGCGCCACCTCCCTGGTGCACCTTGCAAGATCA 480

Y 1113 acttaacctccatagcagcagtagtctcctcaaatcgaaagagcttagactgtcctaac 1172
Db 481 ACTTACTCTCCCTAGCAGAGATAGTCTTCAAAATCGAAAGAGCTTAGACTGTCAACCGC 540

Y 1173 caaagaaggaggaaactgtttatttttagagaagaagcagctgtatattgttaatcaatc 1232
Db 541 TGAAGAGAGGGGGAACCTGTTTATTTTAGGGGAAGAAATCGGTATATATGTTAATCAATC 600

Y 1233 cagaatgtcacctgagaagaatgaaagaatccagatcgatgcaatcaatgtag-agcagagg 1291
Db 601 CGGAATCGTCACTGAGAAAGTTAAAGAAATTCAGATATGAATACAACTANNAAGCANNAG 660

Y 1292 agcttcaaaacacgaacgctggggcctccctcagcaca 1329
Db 661 AGCTTCGAAACACTGGACCTGGGGCTCTCTAGCCAA 698

RESULT 4
LOCUS BE734284 714 bp mRNA linear EST 15-SEP-2000
DEFINITION 601565487F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840572 5',
mRNA sequence.
ACCESSION BE734284
VERSION BE734284.1 GI:10148276
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 714)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLU at: image.llnl.gov
Plate: LNCM530 row: m column: 21
High quality sequence stop: 712.
FEATURES
source
1..714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3840572"
/clone_1id="NIH_MGC_21"
/issue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; site_1: XhoI;
site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 194 a 183 c 137 g 200 t
ORIGIN

Query Match 41.2%; Score 609.8; DB 10; Length 714;
Best Local Similarity 94.3%; Pred. No. 1.5e-151;
Matches 676; Conservative 0; Mismatches 37; Indels 4; Gaps 4;

DB	Y	723	caatctaatgaagcaaacacagctcccaatgcacgaagtgaggtgaacaccccccacagaagt	782
Db	1	CAMTACTACATACACACCAACTCCCAATGCATCAGTGGGTAACTCCCTCCACACAAT	60	
Y	783	agctcgtaccctcaagaaatattcttgcgtgtgtacctagcctacatctgttgaa	842	
Db	61	ACTGTGCTTACCTCAGGAATATTTTGTGTGTGTACTCAGCCTATTC - GTGTTGAA	119	
Y	843	tggctctcaagaattatgttgtctctcattctcttagtgccctctatgacacattacac	902	
Db	120	TGGCTCTTACGATCTATGTGCTTCTCTCATCTTACGTAGGCCCTATGACATCTACAC	179	
Y	903	tgaacaagaattatcaaatcatcgtctacctaagcccccaaaaagatgccattct	962	
Db	180	TGAACAAGATTATACAGTATGCTATCTATCTAAGCCCCGCAACAAAAGAGTACCCATTCT	239	
Y	963	tcccttctgtacagagcagagctgctcgaagcagactaggctacgtgcatgtgcattcac	1022	
Db	240	TCC - TTTGTTATAGAGAGCGAGAGTCTAGTGCACATAGTACTGGCA - TGCGGGATAC	297	
Y	1023	aacctctactcgtctctactacaactatctcaagaataaattgctgataatgaaagct	1082	
Db	298	AACCTTACTCAGTCTCTACTACAACTATCTAAGAACTAATGGGAGCATGGAAGGGT	357	
Y	1083	cactgactccctggtgcaaccttgcaagatacaattactctccctcagcagcagtagtctca	1142	
Db	358	CGCCGACTCCCTGGTGCACCTTCGAAGATCACTTAATCTCCTAGCAACGATGCTCTTCA	417	
Y	1143	aaatcgaagagctcttagactctgtaaccgccaagaagagggagacctgttattttagg	1202	
Db	418	AAATGGAAGAGCTTATAGACTTCTTAACCCCTGAAGAGGGGAGACCTGTTATTTTAGG	477	
Y	1203	agaagaagcgttttttaagttaatcaatccagaattgtcactgtgaaagttaagaagt	1262	
Db	478	GGAAGATCTCTTATATGTTAAATCAATCCGAATCGTCACTGGAAGAAAGTAAAGAAAT	537	
Y	1263	tcgagatcgaaatcaaatgtgagcagagagagcttcaaaaacacgagacgttgggcctct	1322	
Db	538	TGAGATCGAATACAAACGTAGAGCAGAGAGAGCTTGGAAACACTGGAGCCTGGGGCTCCT	597	
Y	1332	cagcaaatgagatgcctgggtctccctctcttaggaacctctagaagctctaatatgtt	1382	
Db	598	CAGCCAAATGGAAGCCCTGGATTCCTCCCTTCTTAAAGACCTCTACGACGATATATTTGCT	657	
Y	1383	actctctcttggaacctgtatcttcaacctctcttgtaagttgtctctccaagat	1439	
Db	658	ACTCCTCTTTGAGACCTGTATCTTTAA - CTCCTGTATTACTTTGTCTTCCAGAAAT	713	
RESULT 5				
LOCUS	BE019603	586 bp	mrna	linear
DEFINITION	bab4f03.y1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:2907101 5'			
	similar to TR:095244 095244 ENVELOPE PROTEIN ;contains Alu			
	repetitive element; , mRNA sequence.			
ACCESSION	BE019603			
VERSION	BE019603.1	GI:8279682		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	NIH-MGC http://mgc.ncl.nih.gov/.			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Washington University Genome Sequencing Center			

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length
36.2%	94.7%	555	0	536.4	9	586
				Pred. No. 5.3e-132		
				Mismatches	31	Indels
					0	Gaps
					0	
703	ctcaccctgtgtaaatattgagcaatctatagacacacacagctcccaatgcatcagtg	762				
1	ctcaccctgtgtaaatattgagcaatctatagacacacacagctcccaatgcatcagtg	762				
763	gtgaacaccccccacacgaaatctgagccacccctcagaataatttttctgtgtgtaac	822				
61	gtgaacaccccccacacgaaatctgagccacccctcagaataatttttctgtgtgtaac	822				
823	tcagccatcatctgtttgaaatgagctctcagaatactatgcttccctcatctcagtg	882				
121	tcagccatcatctgtttgaaatgagctctcagaatactatgcttccctcatctcagtg	882				
883	ccccctatgacatctac	942				
181	ccccctatgacatctac	942				
943	aacaaagaagatcccatctccttctgttcatcagaagagagtgctagcagactagtg	1002				
241	aacaaagaagatcccatctccttctgttcatcagaagagagtgctagcagactagtg	1002				
1003	actgagcatggaagatcac	1062				
301	actgagcatggaagatcac	1062				
1063	aatggtgacatggaac	1122				
361	aatggtgacatggaac	1122				
1123	ctagcagcagtagtcccttcaaaaatcgaagaagcttagactgtctaacgcgaagaag	1182				
421	ctagcagcagtagtcccttcaaaaatcgaagaagcttagactgtctaacgcgaagaag	1182				
1183	ggaacctgttttttttgaagaagaagacgctttttagttaaacaacacagaattgct	1242				
481	ggaacctgttttttttgaagaagaagacgctttttagttaaacaacacagaattgct	1242				
1243	actggaagaagtaagaagaatctcagaatcgaatcagaatcagaatcagaatcaga	1288				
541	actggaagaagtaagaagaatctcagaatcgaatcagaatcagaatcagaatcaga	1288				

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: imgl.jhu.edu/image/html/resources.shtml

Seq primer: -40RP from Gibco

High quality sequence stop: 519.

Location/Qualifiers

1. 586

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2907101"

/clone_lib="NH_MGC_21"

/tissue_type="choriocarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: Placenta; Vector: pOMB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT

168 a 141 c 113 g 164 t

ORIGIN

RESULT 6

BS572445

BS572445

790 bp

mRNA

linear

EST 10-APR-2001

DEFINITION	602593490D1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4720846 5', mRNA sequence.
ACCESSION	BC572445
VERSION	BC572445.1 GI:13580098
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 790)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rgs@bbs-remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LICM1576 row: C column: 23 High quality sequence stop: 624. Location/Qualifiers
FEATURES	1..790
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4720846" /clone_lib="NIH_MGC_79" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site_1: Sfil (ggccatagcgc); Site_2: Sfil (ggccatagcgc); Site_3: Sfil (ggccatagcgc); Site_4: Sfil (ggccatagcgc); Site_5: Sfil (ggccatagcgc); Site_6: Sfil (ggccatagcgc); Site_7: Sfil (ggccatagcgc); Site_8: Sfil (ggccatagcgc); Site_9: Sfil (ggccatagcgc); Site_10: Sfil (ggccatagcgc); Site_11: Sfil (ggccatagcgc); Site_12: Sfil (ggccatagcgc); Site_13: Sfil (ggccatagcgc); Site_14: Sfil (ggccatagcgc); Site_15: Sfil (ggccatagcgc); Site_16: Sfil (ggccatagcgc); Site_17: Sfil (ggccatagcgc); Site_18: Sfil (ggccatagcgc); Site_19: Sfil (ggccatagcgc); Site_20: Sfil (ggccatagcgc); Site_21: Sfil (ggccatagcgc); Site_22: Sfil (ggccatagcgc); Site_23: Sfil (ggccatagcgc); Site_24: Sfil (ggccatagcgc); Site_25: Sfil (ggccatagcgc); Site_26: Sfil (ggccatagcgc); Site_27: Sfil (ggccatagcgc); Site_28: Sfil (ggccatagcgc); Site_29: Sfil (ggccatagcgc); Site_30: Sfil (ggccatagcgc); Site_31: Sfil (ggccatagcgc); Site_32: Sfil (ggccatagcgc); Site_33: Sfil (ggccatagcgc); Site_34: Sfil (ggccatagcgc); Site_35: Sfil (ggccatagcgc); Site_36: Sfil (ggccatagcgc); Site_37: Sfil (ggccatagcgc); Site_38: Sfil (ggccatagcgc); Site_39: Sfil (ggccatagcgc); Site_40: Sfil (ggccatagcgc); Site_41: Sfil (ggccatagcgc); Site_42: Sfil (ggccatagcgc); Site_43: Sfil (ggccatagcgc); Site_44: Sfil (ggccatagcgc); Site_45: Sfil (ggccatagcgc); Site_46: Sfil (ggccatagcgc); Site_47: Sfil (ggccatagcgc); Site_48: Sfil (ggccatagcgc); Site_49: Sfil (ggccatagcgc); Site_50: Sfil (ggccatagcgc); Site_51: Sfil (ggccatagcgc); Site_52: Sfil (ggccatagcgc); Site_53: Sfil (ggccatagcgc); Site_54: Sfil (ggccatagcgc); Site_55: Sfil (ggccatagcgc); Site_56: Sfil (ggccatagcgc); Site_57: Sfil (ggccatagcgc); Site_58: Sfil (ggccatagcgc); Site_59: Sfil (ggccatagcgc); Site_60: Sfil (ggccatagcgc); Site_61: Sfil (ggccatagcgc); Site_62: Sfil (ggccatagcgc); Site_63: Sfil (ggccatagcgc); Site_64: Sfil (ggccatagcgc); Site_65: Sfil (ggccatagcgc); Site_66: Sfil (ggccatagcgc); Site_67: Sfil (ggccatagcgc); Site_68: Sfil (ggccatagcgc); Site_69: Sfil (ggccatagcgc); Site_70: Sfil (ggccatagcgc); Site_71: Sfil (ggccatagcgc); Site_72: Sfil (ggccatagcgc); Site_73: Sfil (ggccatagcgc); Site_74: Sfil (ggccatagcgc); Site_75: Sfil (ggccatagcgc); Site_76: Sfil (ggccatagcgc); Site_77: Sfil (ggccatagcgc); Site_78: Sfil (ggccatagcgc); Site_79: Sfil (ggccatagcgc); Site_80: Sfil (ggccatagcgc); Site_81: Sfil (ggccatagcgc); Site_82: Sfil (ggccatagcgc); Site_83: Sfil (ggccatagcgc); Site_84: Sfil (ggccatagcgc); Site_85: Sfil (ggccatagcgc); Site_86: Sfil (ggccatagcgc); Site_87: Sfil (ggccatagcgc); Site_88: Sfil (ggccatagcgc); Site_89: Sfil (ggccatagcgc); Site_90: Sfil (ggccatagcgc); Site_91: Sfil (ggccatagcgc); Site_92: Sfil (ggccatagcgc); Site_93: Sfil (ggccatagcgc); Site_94: Sfil (ggccatagcgc); Site_95: Sfil (ggccatagcgc); Site_96: Sfil (ggccatagcgc); Site_97: Sfil (ggccatagcgc); Site_98: Sfil (ggccatagcgc); Site_99: Sfil (ggccatagcgc); Site_100: Sfil (ggccatagcgc); Site_101: Sfil (ggccatagcgc); Site_102: Sfil (ggccatagcgc); Site_103: Sfil (ggccatagcgc); Site_104: Sfil (ggccatagcgc); Site_105: Sfil (ggccatagcgc); Site_106: Sfil (ggccatagcgc); Site_107: Sfil (ggccatagcgc); Site_108: Sfil (ggccatagcgc); Site_109: Sfil (ggccatagcgc); Site_110: Sfil (ggccatagcgc); Site_111: Sfil (ggccatagcgc); Site_112: Sfil (ggccatagcgc); Site_113: Sfil (ggccatagcgc); Site_114: Sfil (ggccatagcgc); Site_115: Sfil (ggccatagcgc); Site_116: Sfil (ggccatagcgc); Site_117: Sfil (ggccatagcgc); Site_118: Sfil (ggccatagcgc); Site_119: Sfil (ggccatagcgc); Site_120: Sfil (ggccatagcgc); Site_121: Sfil (ggccatagcgc); Site_122: Sfil (ggccatagcgc); Site_123: Sfil (ggccatagcgc); Site_124: Sfil (ggccatagcgc); Site_125: Sfil (ggccatagcgc); Site_126: Sfil (ggccatagcgc); Site_127: Sfil (ggccatagcgc); Site_128: Sfil (ggccatagcgc); Site_129: Sfil (ggccatagcgc); Site_130: Sfil (ggccatagcgc); Site_131: Sfil (ggccatagcgc); Site_132: Sfil (ggccatagcgc); Site_133: Sfil (ggccatagcgc); Site_134: Sfil (ggccatagcgc); Site_135: Sfil (ggccatagcgc); Site_136: Sfil (ggccatagcgc); Site_137: Sfil (ggccatagcgc); Site_138: Sfil (ggccatagcgc); Site_139: Sfil (ggccatagcgc); Site_140: Sfil (ggccatagcgc); Site_141: Sfil (ggccatagcgc); Site_142: Sfil (ggccatagcgc); Site_143: Sfil (ggccatagcgc); Site_144: Sfil (ggccatagcgc); Site_145: Sfil (ggccatagcgc); Site_146: Sfil (ggccatagcgc); Site_147: Sfil (ggccatagcgc); Site_148: Sfil (ggccatagcgc); Site_149: Sfil (ggccatagcgc); Site_150: Sfil (ggccatagcgc); Site_151: Sfil (ggccatagcgc); Site_152: Sfil (ggccatagcgc); Site_153: Sfil (ggccatagcgc); Site_154: Sfil (ggccatagcgc); Site_155: Sfil (ggccatagcgc); Site_156: Sfil (ggccatagcgc); Site_157: Sfil (ggccatagcgc); Site_158: Sfil (ggccatagcgc); Site_159: Sfil (ggccatagcgc); Site_160: Sfil (ggccatagcgc); Site_161: Sfil (ggccatagcgc); Site_162: Sfil (ggccatagcgc); Site_163: Sfil (ggccatagcgc); Site_164: Sfil (ggccatagcgc); Site_165: Sfil (ggccatagcgc); Site_166: Sfil (ggccatagcgc); Site_167: Sfil (ggccatagcgc); Site_168: Sfil (ggccatagcgc); Site_169: Sfil (ggccatagcgc); Site_170: Sfil (ggccatagcgc); Site_171: Sfil (ggccatagcgc); Site_172: Sfil (ggccatagcgc); Site_173: Sfil (ggccatagcgc); Site_174: Sfil (ggccatagcgc); Site_175: Sfil (ggccatagcgc); Site_176: Sfil (ggccatagcgc); Site_177: Sfil (ggccatagcgc); Site_178: Sfil (ggccatagcgc); Site_179: Sfil (ggccatagcgc); Site_180: Sfil (ggccatagcgc); Site_181: Sfil (ggccatagcgc); Site_182: Sfil (ggccatagcgc); Site_183: Sfil (ggccatagcgc); Site_184: Sfil (ggccatagcgc); Site_185: Sfil (ggccatagcgc); Site_186: Sfil (ggccatagcgc); Site_187: Sfil (ggccatagcgc); Site_188: Sfil (ggccatagcgc); Site_189: Sfil (ggccatagcgc); Site_190: Sfil (ggccatagcgc); Site_191: Sfil (ggccatagcgc); Site_192: Sfil (ggccatagcgc); Site_193: Sfil (ggccatagcgc); Site_194: Sfil (ggccatagcgc); Site_195: Sfil (ggccatagcgc); Site_196: Sfil (ggccatagcgc); Site_197: Sfil (ggccatagcgc); Site_198: Sfil (ggccatagcgc); Site_199: Sfil (ggccatagcgc); Site_200: Sfil (ggccatagcgc); Site_201: Sfil (ggccatagcgc); Site_202: Sfil (ggccatagcgc); Site_203: Sfil (ggccatagcgc); Site_204: Sfil (ggccatagcgc); Site_205: Sfil (ggccatagcgc); Site_206: Sfil (ggccatagcgc); Site_207: Sfil (ggccatagcgc); Site_208: Sfil (ggccatagcgc); Site_209: Sfil (ggccatagcgc); Site_210: Sfil (ggccatagcgc); Site_211: Sfil (ggccatagcgc); Site_212: Sfil (ggccatagcgc); Site_213: Sfil (ggccatagcgc); Site_214: Sfil (ggccatagcgc); Site_215: Sfil (ggccatagcgc); Site_216: Sfil (ggccatagcgc); Site_217: Sfil (ggccatagcgc); Site_218: Sfil (ggccatagcgc); Site_219: Sfil (ggccatagcgc); Site_220: Sfil (ggccatagcgc); Site_2

Y	1102	cttcaagatcaactaactccctcagcagcagtagctccctca-aaatcgaaagagcttga	1160
Y <td>1102</td> <td>cttcaagatcaactaactccctcagcagcagtagctccctca-aaatcgaaagagcttga</td> <td>1160</td>	1102	cttcaagatcaactaactccctcagcagcagtagctccctca-aaatcgaaagagcttga	1160
Db	361	TTGCAGAGTATCAACTTAATCTCCCTAGCACAGTAGTCTTCCACAAATCGAAGACCTTTTGA	420
Y <td>1161</td> <td>cttgcctaacccgcacaaagagggggaacctgttcatlctttagagagaagacgctgttalia</td> <td>1220</td>	1161	cttgcctaacccgcacaaagagggggaacctgttcatlctttagagagaagacgctgttalia	1220
Db	421	CTTGCTAACCGCTGAAGAGAGGGGGAACCTGTTTATTTTATTAGGGGAAGAACTCTGTATTATTA	480
Y <td>1221</td> <td>tgtaaataccacagaaattgttcaactgtagaagaattta-aagaaatccgagatcg--aataca</td> <td>1277</td>	1221	tgtaaataccacagaaattgttcaactgtagaagaattta-aagaaatccgagatcg--aataca	1277
Db	481	TGTTATATCAATCCGGAATCGTACACGAGAAGATTACAGAAGAAATTCGAGATCGCAATACAC	540
Y <td>1278</td> <td>atgttagaccagagagagcttcaaaaacacccgaacgcctggggcctctccagccaatgtagc</td> <td>1337</td>	1278	atgttagaccagagagagcttcaaaaacacccgaacgcctggggcctctccagccaatgtagc	1337
Db	541	ACGTAGACACAGAGGGAGCTTTCGAAMACAGCACACCTGGGGC--TCTCAGCAATGAGATGCC	598
Y <td>1338</td> <td>ctgggtctccctcctctcttagagacctcttagc-agctcctaataltgttactcctcttgagc</td> <td>1396</td>	1338	ctgggtctccctcctctcttagagacctcttagc-agctcctaataltgttactcctcttgagc	1396
Db	599	--TGATATTCCTCCCTCTTAGGACCTCTAGCAAGCTATACATATGCTACTCTCTC--TTGAG	655
Y <td>1397</td> <td>cctgtatctttaaactcctctgttgaattgtctctcttcagagaatgaaagctgtaaagct</td> <td>1454</td>	1397	cctgtatctttaaactcctctgttgaattgtctctcttcagagaatgaaagctgtaaagct	1454
Db	656	CCTGTATCTTTAACTCCTCTGTCAACTGAGTCTCTTCAGAAATCGAAGGTATAACCT	713
RESULT 7			
AA776439/c			
LOCUS	AA776439	427 bp	MRNA
DEFINITION	zj50h06.s1 Soares_fetal_liver_spleen_lNfLS_S1 Homo sapiens cDNA		
	clone IMAGE:553755 3' similar to SW:ENV_SMRVH P21412 ENV		
ACCESSION	AA776439		
VERSION	AA776439.1 GI:2835773		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 427)		
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,		
	Krizman,D., Kucaba,T., Lacey,M., Le,N., Lennon,G., Marra,M., Martin		
	J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
	White,Y., Wylie,T., Waterston,R. and Wilson,R.		
TITLE	WashU-NCI human EST project		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LLNL; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Seq primer: -40m13 fwd. ET from Amersham		
	High quality sequence stop: 410.		
FEATURES	Location/Qualifiers		
SOURCE	1..427		
	/organism="Homo sapiens"		
	/db_xref="GDB:1139011"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:453755"		
	/clone_lib="Soares_fetal_liver_spleen_lNfLS_S1"		
	/sex="male"		
	/dev_stage="20 week post conception fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: Liver and Spleen; Vector: pTZ193D (Pharmacia)		
	with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;		
	This is a subtracted version of the original Soares fetal		
	liver spleen lNfLS library. 1st strand cDNA was primed		
	with a Pac I - oligo(dT) primer [5'		
	AACGCGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'].		
	double-stranded cDNA was ligated to Eco RI adaptors		

QY 733 gacacaaacagcctcccaatgcatcagtggtgtaacacccctcccaacgaatagctgccta 792
 Db 121 TACACAACCACTCCCAATGCATCAGTGGTGAATCTCTCCACACAATAATGCTGCTTA 180
 QY 793 cccctagggaatatttttctgctgctgtaacctgaacctatattgttgaatgctcttca 852
 Db 181 CCTCAGGAATATTTTGTCTGTGTGACCTGACCTATCGTTGTGAATGGCTCTTCA 240
 QY 853 gaactatgctctccctcattcttagtgccccctatgacacatcacaactgaacaagat 912
 Db 241 GATTCATATGCTCTCTCTCTATCTTAGTCCCTCATGACATCTACACTGACAAGAT 300
 QY 913 ttaacaatcaatgctgtaacctaa-gccccacaacaaagaatgaccca-ttctctcttgg 970
 Db 301 TTATACAGTATGTATCATATATTAAGCCGCCACAAAAGAGTACCATTCTCTCTTTG 360
 QY 971 ttat-cagagcagatgctcaggcagatcagtagctagctggtcagatgatacaacctc 1029
 Db 361 TTATAGGAGACAGGGGTGTGGGTGATGGGTGANTGGCGTGGCGTATCACACCTTT 420
 QY 1030 acccagctc-tactacaactatcctcaagaataatgg 1067
 Db 421 ATTCACTTTTATTACAAATTTCTCAGGANTATATGG 459

RESULT 11 608 bp DNA linear GSS 14-JUL-1999
 LOCUS AO725469/c HS_5403.AL.E02.T7A RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-979 Col-3 Row-1, DNA sequence.
 ACCESSION AO725469
 VERSION AO725469.1 GI:5485138
 KEYWORDS GSS.

SOURCE

human.

Homo sapiens

REFERENCE

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 1 (bases 1 to 608) Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 Kellier, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

FEATURES

source

Class: BAC ends
 High quality sequence stop: 608.
 Location/Qualifiers
 1..608
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-979 Col-3 Row-1"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"
 BASE COUNT 165 a 114 c 145 g 173 t 11 others
 ORIGIN

Query Match 24.1%; Score 356.4; DB 12; Length 608;
 Best Local Similarity 88.3%; Pred. No. 4,8e-84;
 Matches 406; Conservative 0; Mismatches 52; Indels 2; Gaps 2;

QY 1022 caaccctctactgctctctctacaaactatcagaagaataatggtgacatggaacagg 1081
 Db 608 CAACCTCTACTCAGTTTACT-CAACATATCTCAAGATTAATGGGTGACATGATGATG 550
 QY 1082 tcaactactccctcgtcaccctgcaagatcaactaactcctcagcagcagtagctctc 1141
 Db 549 TCACNGACTCCCTGTCACCTTCACAGATCTCAAGATCTTAACTCTTAAAGCAGATG 490
 QY 1142 aaaaatcgaagagctttagactgctaaacgcacaaagaaggggaacctgttaatttag 1201
 Db 489 -AAATCGAAGAGCCTTAAAGACTTCTTAAGTCTCAAGAAAGAGAGGAAATCTGTTT 431
 QY 1202 gagaagaagcgtgttatattgtaataatcaccgaattgctactgagaagtaagaaa 1261
 Db 430 TGAAGAAGATGCTGTATTAAGTTAACTGAATCTGAATCGCACGAANAAGTTAAGAAA 371
 QY 1262 ttccagatcgaataacataatgtagagcagagagcttcaaaacacagcagctgggctcc 1321
 Db 370 TTCAAGATGCAATATCAACATATAGAGAGAGAGGTTTCANANACACAGACCTGGAG 311
 QY 1322 tcaagcaatgagatgacctgggtctccctctcttaggaacctcagacagctcaattgt 1381
 Db 310 TCAGCCAAATGAGATGCCCTGATCTCCCTCTTACGACCTTACGACATATATCTG 251
 QY 1382 tactcctcttggaccctctatcctttaaactctgtttaggttgcctctcagaagatt 1441
 Db 250 TACTCTCTTTGAGACCTGATCTTAACTCTTTTAAAGTTGCTCTTCAGAAATTTG 191
 QY 1442 aagctgtaaagctacagatgctcttcaaaatggaaccca 1481
 Db 190 AAGCTGTAAGCTAACAAATGGCTCTTCAATATGAGAGCCCA 151

RESULT 12 411 bp mRNA linear EST 24-JAN-2001
 LOCUS BG004254/c RC6-GN0070-301100-014-C06 GN0070 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG004254
 ACCESSION BG004254
 VERSION BG004254.1 GI:12445242
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 411)

REFERENCE

AUTHORS

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPSP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&tl2=RC6-GN0070-30100-014-G06&f3=2000-11-30&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 21
High quality sequence stop: 411.
Location/Qualifiers

FEATURES
source
1..411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="GN0070"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site:1: SmaI
; Site:2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 103 a 94 c 96 g 118 t
ORIGIN

Query Match 24.0%; Score 355.2; DB 10; Length 411;
Best Local Similarity 94.1%; Pred. No. 9.3e-84;
Matches 369; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 1005 tggcattggcagatcaaacaccttactcagttcttactcaaacatctcagaagaataa 1064
Db 411 TGGCATTGGCGGTATCACACACCTTACTACTACTACACAGCTACTCAAGACTAA 352
Oy 1065 tggtagacatggaacagctgactgactccctgtgctcactgcaagaataaactccct 1124
Db 351 TGGGACATGGAGACGGGTGCGCGACTCCCTGTCACCTTGCACAGTCAACTTAACCTCT 292
Oy 1125 agcagcagtagtccctcaaatcgaaagactttagactgtcaaccgcaaaagaagag 1184
Db 291 AGCAGCAGTAGTCCCTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGGGGG 232
Oy 1185 aacctgttatttttagaagaagaacgctgttattatgttaataatccagaattgtcac 1244
Db 231 AACCTGTTATTATTTTAAAGGGAAGATGCTGTTATTATGTTATCAATCCGGAATGTCAC 172
Oy 1245 tgaagaagttaagaagaatctcagatcgaaatacaatgtagagaagaagaagctcaaaac 1304
Db 171 TGAGAAATTTAAAGAAATTCAGATCGAATACAGCTGAGAGAGAGACTTCGAAACAC 112
Oy 1305 cgaacgctggggcctcctcagaacatgagatgagctggttctccctctttagaactct 1364
Db 111 TGGACCTTGGGGCCTCCTCAGCCAAATGAGTGGCTGGAATTCCTCCCTATTATGAGACTTCT 52
Oy 1365 agcagcttaatatgttactcctctttggac 1396
Db 51 AGCAGCTATATATGCTACTCTCTTTGGAC 20

RESULT 13
BI052569/c 382 bp mRNA linear EST 15-JUN-2001
LOCUS B1052569
DEFINITION RCS-GN0282-220101-011-G04 GN0282 Homo sapiens CDNA, mRNA sequence.
ACCESSION B1052569
VERSION B1052569.1 GI:14460099
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 382)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

JOURNAL TITLE
MEDLINE
COMMENT
'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPSP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&tl2=RC5-GN0282-
220101-011-G04&f3=2001-01-22&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 381.
Location/Qualifiers

FEATURES
source
1..382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="GN0282"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site:1: SmaI
; Site:2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 92 a 55 c 121 g 114 t
ORIGIN
Query Match 23.6%; Score 349.6; DB 10; Length 382;
Best Local Similarity 95.0%; Pred. No. 2.8e-82;
Matches 361; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

Oy 411 ccgggagacatgagccctcagccctcacaagaagacttctctcacaactacatgaac 470
Db 380 CCGGATACATGACACCTCTAGCCCTCAAAAGACTAGTCTCAAAATCTACGAAAC 321
Oy 471 cctcgtaccatctgctgctggtgagctatttaataacacccctcactggtcctcaga 530
Db 320 CCTCGTACCCCTACTCGCTGTGTAAGCCTATTATACACCCCTACCTGGGCTCATTA 261
Oy 531 ggtccagcccaaacctcactactgttggatgtgctccctcgcactcaggcata 590
Db 260 GGTCTCGGCCCAAAACCTACTACTGTGGATGATGCTCCCTTGMACTTAGGCCATA 201
Oy 591 cattcaatccctgttctcgtgaacaaatggaacaactcagcacagaataaacacacttc 650
Db 200 TGTTCATTCCTGTACCTGAGACAAATGGAACAATTCACAGAAATTAACACACACTTC 141
Oy 651 cgtttagagagactctgttccaatctggaataaaccatctactcctaactcactg 710
Db 140 CGTTTATGAGGACCTCTGTTTCCAAATCTGGAATTAACCCCTACTCAAACTCACTG 81
Oy 711 tgaataatttagacaatacatagacaacaacagcctcccaatgcatcaggtggtaacac 770
Db 80 TGTAAATTTAGCATACTACTACACAAACCACTCCCAATGCATCAGTGGGTAACTCC 21
Oy 771 tccacacgaatagctcgc 790
Db 20 TCCACACCAATATGCTGCC 1

RESULT 14
BI056595/c 376 bp mRNA linear EST 15-JUN-2001
LOCUS B1056595

DEFINITION	RC5-GN0282-210201-013-604 GN0282 Homo sapiens cDNA, mRNA sequence.
ACCESSION	B1056595
VERSION	B1056595.1 GI:14464125
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P., Goldman, G.H., Carvalho, A.F., Matsumura, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.U. and Simpson, A.U.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20020663
COMMENT	Contact: Simpson A.U.J. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-3704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC5&t2=RC5-GN0282 210201-013-G04&t3=2001-02-21&t4=1) Seq primer: puc 18 forward High quality sequence stop: 376. Location/Qualifiers 1..376 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="GN0282" /dev_stage="Adult" /note="Organ: placenta.normal; Vector: puc18; Site_1: Small ; Site_2: Small; A mini-library was made by cloning products derived from ORESes PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
FEATURES	
source	
BASE COUNT	90 a 52 c 120 g 114 t
ORIGIN	
Query Match	23.6%; Score 349.4; DB 10; Length 376;
Best Local Similarity	95.7%; Pred. No. 3.2e-82;
Matches 359; Conservative	0; Mismatches 16; Indels 0; Gaps 0;
OY	418 catagacccttagccctcacaagaagactagtctctcaaaacatcatgaaaccttcgt 477
Db	376 CATGCACTTACGCCCTTACAAAGACTAGATCTCTCAAACATCAGTAAGAACCTCCGT 317
OY	478 acccctactgcgctgtgaagcatttaataccacctccatcgctcatatgaagttcca 537
Db	316 ACCCATACTCGCTTGTAAGCCATTATTAATACCCTCTCACGGCTCCATGAGGTCTCG 257
OY	538 gcccaaaaacctaactatgttgaatgtgcctcccctggcaacttcaggccatatctca 597
Db	256 GCCCAAAACCCCTACACTGTTGATGATGCGTCCCCCTGAAGCTTCAGGCACAATGTTTCA 197
OY	598 atccctgttccctgaacaatlygaacaacttcagcacagaataaacaacacttcgltta 657
Db	196 ATCCCTGTACCGTAGAACAATGACAACTTCAGCACAGAATAAACAACACACTTCCTTTA 137
OY	658 gtgagacctcttgttctcaatcttgnaaataccatlaactcataactcaactcgtgttaa 717
Db	136 GTAGGACCTCTTGTTTCCAATCTGGAATTAACCCCTACTCAAAACCTCAACCTGTTTAAA 77

Oy	718	ttagcaatacatatagaccacacgcgtcccatgatcagtggtaaaccttcocaca	777
Dd	76	ttttagcaaatgactgacacacgaacctcccatgattcggttggttaccttcocaca	17
Oy	778	cgaaatgctgccta	792
Dd	16	CAAAATGCTGCCTA	2
RESULT	15	B1052559	373 bp mRNA linear EST 15-JUN-2001
LOCUS	B1052559/C	RC5-GN0282-220101-011-A04 GN0282 Homo sapiens cDNA, mRNA sequence.	
DEFINITION	B1052559		
ACCESSION	B1052559.1	GI:14460089	
VERSION	EST.		
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 373) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balz,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.U.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
TITLE	Contact: Simpson A.U.J. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel.: +55-11-7704922 Fax: +55-11-7707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&t2=RC5-GN0282- 220101-011-A04&t3=2001-01-22&t4=1) Seq primer: puc 18 forward High quality sequence stop: 373. Location/Qualifiers 1..373 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="GN0282" /dev_stage="Adult"		
FEATURES	Source	/note="Organ: Placenta-normal; Vector: puc18; Site_1: Small ; Site_2: Small; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	89 a 52 c 120 g 112 t		
ORIGIN			
Query Match	23.5%	Score 347.4;	DB 10; Length 373;
Best Local Similarity	95.7%;	Fred. No. 1.le-81;	
Matches 357; Conservative	0;	Mismatches 16;	Indels 0; Gaps 0;
Oy	418	catacacccccaggcccctacaagaagactgcttcctcaaaactatcagaaacctcgt	477
Dd	373	CATTGGCACCCTCTAGCCCTTACAAGAAGACTAGATCTCTCAAAAATCATGAACCTCCGT	314
Oy	478	accatactgcctctgtagcctattataaccacccctcactcggctccatgaggtctca	537

Db 1168 catgcaatactcatattgacacgaaaaatgaattaatctaatgtctgaagactt 1227

Oy	301	ggaagcaactctctgttgactactctacacccaacagatgtctcgatggggttgaaat	360
Db	1228	ggagatccctgctctgttgactactctacacccaacctgcatctcgatggggttgaaat	1287
Oy	361	caagatccagagcaagagaaaacagtaaaagagcaatctcccaactgacccgggagat	420
Db	1288	caagatccagagcaagagaaaacagttaaaagaagtatctcccaactcaccgggtacat	1347
Oy	421	agcccccttagcccttcacaaagacagtctctctcaaaactcatgtaaaacctccgtac	480
Db	1348	ggcccccttagcccttcacaaagacagtctctctcaaaactcatgtaaaacctccgtac	1407
Oy	481	catatctgcctgtgagccatttaatacacaacctacccgtctccatgagttccac	540
Db	1408	catatctgcctgtgagccatttaatacacaacctacccgtctccatgagttccgc	1467
Oy	541	caaaaaccttaactgtgtgagtgcctccctgcactctcaggccaacatttaact	600
Db	1468	caaaaaccttaactgtgtgatagtccctccctgaaactcaggccaatgttttaact	1537
Oy	601	ccgttctctgaacatgagaaacaacttcagacagaaataaacacactccgtttta	660
Db	1588	ccgttactctgaacatgagaaacaacttcagacagaaataaacacactccgtttta	1587
Oy	661	ggacctctgtttccaaactctggaataaacccaatacctcaaacctcacctgtgtaaatt	720
Db	1588	ggacctctgtttccaaactctggaataaacccaatacctcaaacctcacctgtgtaaatt	1647
Oy	721	agcaatactctagagcaacaacagctcccaatgatacaggttggtgtaaacctccaca	780
Db	1648	agcaatactctatacaacaacaacccccaagatacaggttggtgtaaacctcccca	1707
Oy	781	atagtccgcctacccctccaggaataattcttgctgtgtgacccagccatcatgtttg	840
Db	1708	atagtctgcctacccctccaggaataattttgtctgtgtgacccagccatcatgtttg	1767
Oy	841	aatgtgcctctcagaaattatagtctctccctcaattcttagtgtccccctatgacatctac	900
Db	1768	aatgtgcctctcagaaattatagtctctccctcaattcttagtgtccccctatgacatctac	1827
Oy	901	actgaacaagatctatacaatctgcgtgacccaagaacccccaacaagaagtaacct	960
Db	1828	actgaacaagatctatacaatctgcatactataagccccgacacaaaaggtaacctat	1887
Oy	961	cttctctcttctatcagagcagagagtgcacgacagactagtaactgtgcattgcagatc	1020
Db	1888	cttctctcttctatagagagagagagtgcactgagctagacgtgacattgtgcgttatc	1947
Oy	1021	acaaacctctacagtctctactcaacaactatccaaagaaataatgtgtgactgtgaacag	1080
Db	1948	acaaacctctactcagttctctactcaacaactatctcaagaactaatgtgtgagaaagg	2007
Oy	1081	gtcacgtactccctgtgcaccttgcagaatacacttaactctcctatgcagcagtaagtcct	1140
Db	2008	gtcgcgcgactccctgtgcacacttgcagaatacacttaactctcctatgcagcagtaagtcct	2067
Oy	1141	caaaattcgaagagctttaagctgtgtaacccgtgaaaggggggaacctgtttatttta	1200
Db	2068	caaaattcgaagagactttaagctgtgtaacccgtgaaaggggggaacctgtttatttta	2127
Oy	1201	ggaaagaaagacgctgttctatctgtctaatccaatccagaaatgtctaacctgaaagttaa	1260
Db	2128	ggagaaagaaatgctgttctatctgttataatccaatccagaaatgtctaacctgaaagttaa	2187
Oy	1261	atccgagatccgaatacaatgttagagcagagagctcaaaacacggaacgctgggctc	1320
Db	2188	atccgagatccgaatacaacgttagagcagagagcttcgaaacaacctgagccctgggctc	2247
Oy	1321	ctcagcgaatggagatgccctgtggtctccctctcttagagccctctgagagccttaattg	1380
Db	2248	ctcagcgaatggagatgccctgtgattccctctctcttagagccctctgagagccttaattg	2307

```

OY 1381 ttaccctcttggagcccggtatctttaaaccctgttaagtttgcctctccagaatt 1440
Db 2308 ctaccctcttggagcccggtatctttaaaccctgttaagtttgcctctccagaattc 2367
OY 1441 gaagctgttaagctacagatggtcttacaatatggaaccca 1481
Db 2368 gaagctgttaaacatacaaatgagcccaagatgcaatcca 2408

RESULT 2
US-08-686-878A-48
; Sequence 48, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavalley, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-686-878A-48

Query Match 11.7%; Score 173.8; DB 1; Length 542;
Best Local Similarity 91.5%; Pred. No. 1.4e-46;
Matches 184; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY 1 atggcctccctctacacactttctcttcttaactgtctctcttaaccccccttcgctcact 60
Db 342 atggccctccctctacacattttctcttcttaactgtctcttcttaaccccccttcgctcact 401
OY 61 gcaecctccatcgtctgctgtacaacacagtagctccctcttacaagaagttctatgaaga 120
Db 402 gcacccctccctctacacattttctcttcttaactgtctcttcttaaccccccttcgctcact 461
OY 121 acggagctctccgggaatatatgatgccccatcataatagagtttctcaagggaaccc 180
Db 462 atgcagcgtcccggaatatatgatgccccatcgtatagagcttttcttaagggaaccccc 521

```


D _b	646 GTAGAGATGGCACCTGGCATTTACAGAAAGGCGCTTGTGATTACAGACAATGCCCTTCA	705
Oy	1300 aaacgcgaagctgtagggcccccctaagcgaatlgatgccccagggtttccccccttlaaga	1350
D _b	706 AACTCTTAACCA-----CCTCTGGAGTAGTGGAACAATGCCTTTCTTCATTTC TAGST	759
Oy	1360 cctctagcagctcttaatatgtttaactccttcctlttgagaccctgtaaccttaacctccttgtt	1410
D _b	760 CCCATGGCAGGCATCTTGCTGTACTACCTTTGGGCCCCGTATTTTTAAGCTTCCTGTGC	819
Oy	1420 aaagtctgcctcttcagaattgaagcgtlaaaagctcaagatgctcttacaatlgaacc	1470
D _b	820 AAATTTGTTGTTCTCTAGATCGAAAGCCATCAACTACAGATGCTCTTTACAATAATGGAACC	879
Oy	1480 ca 1481	
D _b	880 CA 881	

SULT 5
US-08-691-563C-46
Sequence 46, Application US/08691563C

1 GENERAL INFORMATION:
 2 APPLICANT: Herve PERRON
 3 APPLICANT: Frederic BESEME
 4 APPLICANT: Frederic BEDIN
 5 APPLICANT: Glaucia PARANHOS-BACCALA
 6 APPLICANT: Florence KOMURIAN-PRADEL
 7 APPLICANT: Colette JOLIVET
 8 APPLICANT: Bernard MANDRAND
 9 TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
 10 TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
 11 TITLE OF INVENTION: THERAPEUTIC PURPOSES
 12 NUMBER OF SEQUENCES: 92
 13 CORRESPONDENCE ADDRESS:

ADDRESS: Oliff & Berridge
STREET: 700 South Washington Street, Suite 3000
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C

```

1 BILLING DATE: 02-AUG-1996
2
3 ATTORNEY/AGENT INFORMATION:
4
5     NAME: Berridge, William P.
6
7     REGISTRATION NUMBER: 30,024
8
9     REFERENCE/DOCKET NUMBER: WPB 385688
10
11 TELECOMMUNICATION INFORMATION:
12
13     TELEPHONE: 703-836-6400
14
15     TELEFAX: 703-836-2787
16
17 INFORMATION FOR SEQ ID NO: 46:
18
19     SEQUENCE CHARACTERISTICS:
20
21         LENGTH: 1859 base pairs
22
23         TYPE: nucleic acid
24
25     STRANDEDNESS: single
26
27     TOPOLOGY: linear
28
29     MOLECULE TYPE: CDNA
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

Query Match	8.9%	Score 131.2	DB 3	Length 1859
Best Local Similarity	57.6%	Pred. NO. 2e-32		
Matches 257, Conservative	0	Mismatches 183	Indels 6	Gaps 1

by 1036 ttctactacaactatctcaagaataaatggtgacatggaacaggtcactgactccctg 1095

Db	1092	TACTACACACACTCTCAAGGATTTCTCAGACAGTTTGCCAGAAATAATGATATCTATC	1151
Qy	1096	gtcaccttgcagaagatcaacttaacccccctagcagcagtaglccctcaaatcgaagagct	1155
Db	1152	CTTACTCTACAAATCCCAATAATAGACTCTTTGGCAGCAGTACTCTCCAAAACCGTCAGAGC	1211
Qy	1156	ttgaacttgcctaacccgcacaaagaggggagactcttatttctttagtggaagaagcgcgt	1215
Db	1212	CTTACACCTCCCTCANTGCTGTGAGAAAGAGACCTCTGCACCTTTTAAAGGAAAGAGTGGT	1271
Qy	1216	tattatgttaatcaatccagaattgtcacgtgaagaagttaaagaattcgcagatcgaaat	1275
Db	1272	CTTTACACTTACCCAGTCAAGGATAGTATGATGCTGCCCGGCATTTTACAGAAAAAGCT	1331
Qy	1276	caatgttagcagcagagagcttccaacacacgaagcgtgggcctctcagcaalbgatg	1335
Db	1332	TCTGAATTCAGACACAGCCCTTTCAAATTC-----CTATACACACCTCTGGAGTTGGCA	1385
Qy	1336	ccctgggtctcccccctcttagaacctctagcagcctcaatatgttctactcccttggga	1395
Db	1386	ACATGAGTTTCTCCCTTTCTATCTCCCATGGCGCCANTCTCTATTCAGCGCCCTTGGG	1445
Qy	1396	ccctgatactttaaaccctctgttaagtttgcctctctccagatctgaagcctglaaagcta	1455
Db	1446	CCCGTATTTTAACTCCCTTTGTCAAAATTTGTTCTTTAGGATGAGGACATCAAGCTA	1505
Qy	1456	cagaigtctacaatatgaaacca 1481	
Db	1506	CAGATGGCTTTACAAATGAAACCCCA 1531	

RESULT 6
US-08-007-282B-1
; Sequence 1, Application US/08007282B

Patent No. 5403582
GENERAL INFORMATION:
APPLICANT: NAZERIAN, KEYVAN
APPLICANT: CALVERT, JAY G.
APPLICANT: WITTER, RICHARD L.
APPLICANT: YAMAGIDA, NOBORU
TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road Suite 500 East
CITY: Falls Church
STATE: VA

1 COUNTRY: USA
2 ZIP: 22042
3
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: FLOPPY disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8 SOFTWARE: PatentIn Release #1.0, Version #1.25
9
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/007,262B
12 FILING DATE: 1993/0121
13
14 CLASSIFICATION: 424
15
16 ATTORNEY/AGENT INFORMATION:
17 NAME: MURPHY JR., GERALD M.
18 REGISTRATION NUMBER: 28,977
19 REFERENCE/DOCKET NUMBER: 1644-104P
20
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: 703-205-8000
23
24 TELEFAX: 703-205-8050

```

? TELE: 248345
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1704 base pairs
? TYPE: NUCLEIC ACID
?
```

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
US-08-007-282B-1

Query Match 7.9%: Score 116.8; DB 1; Length 1704;
Best Local Similarity 51.5%: Pred. No. 9.2e-28;
Matches 268; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

Qy 932 ctaagccccacaacaaagagtagccatctctcttctgtatcagagcagagtagctag 991
Db 1178 ctAGCGGcATTAAGAGGGcAGTCCAGTTATCCCGCTGTTGGGTCTAGGAGATTTCAG 1237
Qy 992 gcagactagtagctggcattggcagatcacacacctactcagtlctactacaacta 1051
Db 1238 GGGCTACACTTGTGGTGAACGGGGCTTGGGGTCTCCGTTCCACTTATCACAAGCTCT 1297
Qy 1052 ctcaagaataatagtagcatgagacagtagcactgactctgtctacacttgcaagatc 1111
Db 1298 CTATTCATTTGATTGAGATGTCCAGGCTCTTTCAGGACATCAATGACTACAGAC 1357
Qy 1112 aacttaactccctagcagcagtagctctcaaaatcgaagagcttagactgtccacag 1171
Db 1358 AGATTGACTCCCTGGCTGAGGTGTCTTACAAATAGAGAGGGTTAGACCTATTGACTG 1417
Qy 1172 ccaaaagaggaggaacctgttatttttagaggaagaacgctgtattatgttaacat 1231
Db 1418 CCGAACAAGAGGAGATATGTCTCCGACCTCCAGAGAGAGTGTGTTTACCGCTACAAGT 1477
Qy 1212 ccaagaattgtagcaggaagttaaagaattcgagatcgaataaagttagagcagag 1291
Db 1478 CGGATGTCAGTGACAGATCGAAATCTCCAGAGAGACCTTATCGAGAGAAAAAGTg 1537
Qy 1292 agctcaaaacacgcaagcgtgggctctccacgaatgtagcctgtgtctccct 1351
Db 1538 CACTGTACGACAAACCCCTGTGGAGCGGCTTGAACGGCTTCCTTCATATTGTGACCT 1597
Qy 1352 tcttagagaccttagcagcttaatatgttactcctcttggagccgtatcttaacc 1411
Db 1598 TGTTAGGCCCCCTTGTGGCTCATATTGTTCCATACCTCGGCCCGGCGCATTAAGAGA 1657
Qy 1412 tccctgttaagttgtctctccagaattgaagctgttaa 1451
Db 1658 CCTGTACTCGCATTTATACATGACAAATTCAGCGAGTAA 1697

RESULT 7

US-09-078-294-4
Sequence 4, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sarf, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078, 294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 80246
TYPE: DNA
ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 7.8%: Score 114.8; DB 4; Length 80246;
Best Local Similarity 55.7%: Pred. No. 3.8e-26;
Matches 263; Conservative 0; Mismatches 202; Indels 7; Gaps 2;

Qy 1010 ttggcagtatccacacctctactcagttctactacaactatctcaagaataatggtg 1069
Db 58509 ttaacagcgtacacatglttaltccactactcactcaactctctcaagaatttctcagaca 58568
Qy 1070 acatggaacaggtcacctactcctctgctacacttgcaagatcaactactcctcagag 1129
Db 58569 gttgcaaaaaaagaacgaatctgtctctactctactacaatcccaatagactcttggcag 58628
Qy 1130 cagtagtccctcaaaaatcgaagagctttagactgtctaacgccaagaaggggaacct 1189
Db 58629 cagtagctctccaaaacgcgtgagcctagactctctactctgtgagaaggaagattct 58688
Qy 1190 gttattttaagagaagaacctgttattatgtaataccaacagaattgtcactgag 1249
Db 58689 gac-ttcttaagggttagagtggttcttactaaccacagcagggaataatagagaa 58747
Qy 1250 aagttaaagaatctgaagatcgaatacaatgtagagcagagaggtcnaaacacggaac 1309
Db 58748 ccaccagtggtttacaggaagaagctctcgaatacagacaatgaccttcaacttata 58807
Qy 1310 gctggggcctctcagcaatgtagtccctgggtctccctctttagagactctagcag 1369
Db 58808 ccaa-----cctctggagtggtggcagacatggtctctcccttcttagtcttgagcag 58861
Qy 1370 ctctaattgttactcctcttggacacctgtatcttcaactcctgtttaggttgcct 1429
Db 58862 ccacttgctaatagtcgacttggccctgtatcttcaactccttggccaatttgcct 58921
Qy 1430 ctccagaattgaagctgtaagctacagatggtcttcaaatggaaccca 1481
Db 58922 cctctagtagtgagggcatcaagctacagatgattcttcaaatgtgaaccca 58973

RESULT 8

US-09-011-745-1
Sequence 1, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiko
APPLICANT: Cosset, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2518
TYPE: DNA
ORGANISM: RD114
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)
OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

Query Match 7.7%: Score 114.2; DB 4; Length 2518;
Best Local Similarity 55.0%: Pred. No. 8e-27;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 1045 aaactatcctaagaataatgtagacatgaggaaggtacagactcctctgtctaccttg 1104
Db 1834 aaattatcccatcagttatcatctgtagtccaggtcttaccggtacacataagaattta 1893
Qy 1105 caagatcaacttaactcctcagcagcagtagccttcaaaatcgaagaagcctttagacttg 1164

```

Db 1894 caagaccagctgactcgttagctgaagtagtctccaaatagaggagcggacccta 1953
Oy 1165 ctaccgccaagaagagggaacctgttatttttagaagaagacgcttaattgt 1224
Db 1954 ctacggcgcaagaagaaggaattttagccttaagaagaagacgcttttttgc 2013
Oy 1225 aatcaatcagaattgtcactgagaagaattgaagaatcagatcgatatacaatgtaga 1284
Db 2014 aacaagtcaggaattgtgagaacaataaagaacccctcaagaagaattacaagaacgc 2073
Oy 1285 gacgaaggagcttcaaaaccgaacgctgggctcctcagccaatgagtgccctggt 1344
Db 2074 agggaaagcctggcaaccacccctctgacccggctgcaaggccttctccgtaacc 2133
Oy 1345 ctcccctcttgagacctctagacgtcctaattgttactcctcttgagccctgtac 1404
Db 2134 ctaccctctctggaccctactcaaccctcctactcaatacctatgggccaatgct 2193
Oy 1405 tttaacctctgttaagtgtgtctctcctccagaattgaagctgttaa 1451
Db 2194 ttacgctcctcatggccttcaatgataagacttaactgtgtaca 2240

```

```

RESULT 9
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Weiss, Robin A.
; APPLICANT: Collins, Mary KL
; APPLICANT: Takeuchi, Yasuhiko
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3611)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3612)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3613)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3614)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3799)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3800)

```

```

; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3801)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3802)
; OTHER INFORMATION: n is any nucleotide
; US-09-011-745-8

```

```

Query Match 7.7%; Score 114.2; DB 4; Length 5865;
Best Local Similarity 55.0%; Pred. No. 1.3e-26;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

```

```

Oy 1045 aaactatccagaataatagtgtgacataggaacaggtacacgacctgtgtacactgt 1104
Db 2692 aaattatccatcagttatattctgacatgctcgttaccatataagaattta 2751
Oy 1105 caagatcaactaacctcctcagcagtagtccctcaaatcgaagacttaagactgt 1164
Db 2752 caagaccagtagactcgttagctgaagtagtctcacaataagagagggactgaccta 2811
Oy 1165 ctaccgccaagaagagggaacctgttatttttagagaagaagacgctgtattatgt 1224
Db 2812 ctacggcgcaagaagaaggaattgttagcctcaagaagaatgctgttttagct 2871
Oy 1225 aatcaatcagaattgtgtcctgagaaggttaagaatcagatcgatgataatgtaga 1284
Db 2872 aacaagtcaggaattgtgagaacaataaagaacccctcaagaagaattacaagaacgc 2931
Oy 1285 gacgaaggagcttcaaaaccggaacgctgggctcctcagccaatgagtgccctggt 1344
Db 2932 agggaaagcctggcaaccacccctctgacccggctgcaaggccttcttcgtaacc 2991
Oy 1345 ctcccctcttgagacctctagacgtcctaataatgttactcctcttgagccctgtac 1404
Db 2992 ctaccctctctggaccctactcaaccctcctactcaatacctatgggccaatgct 3051
Oy 1405 tttaacctctgttaagtgtgtctctcctccagaattgaagcgtgttaa 1451
Db 3052 ttacgctcctcatggccttcaatgataagacttaactgtgtaca 3098

```

```

RESULT 10
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desirée
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies COI
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contlg
; US-09-078-294-3

```

```

Query Match 6.1%; Score 90.8; DB 4; Length 80595;
Best Local Similarity 57.5%; Pred. No. 2.4e-18;
Matches 206; Conservative 0; Mismatches 142; Indels 10; Gaps 2;
Oy 1124 tagcagcagtagtctcctcaaaatcgaagagctttagactgttaacggccaagaaggg 1183
Db 58859 tggcagcagtagtctcctcaaaacgctgagccttagactctcaatgtttgagaagaag-- 58916

```


RESULT 11
US-09-011-745-9
Sequence 9, Application US/09011745
Patent No. 6155715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiko
APPLICANT: Cossetl, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9

```

? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Portion of
? OTHER INFORMATION: construct
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3910)
? OTHER INFORMATION: n is any nucleotide
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3911)
? OTHER INFORMATION: n is any nucleotide
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3912)
? OTHER INFORMATION: n is any nucleotide
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (3913)
? OTHER INFORMATION: n is any nucleotide
? OS-09-011-745-9

```

Query Match	4.7%;	Score 70;	DB 4;	Length 3925;
Best Local Similarity	49.3%;	Pred. No. 2.4e-12;		
Matches 224;	Conservative	0;	Mismatches 215;	Indels 15;
				Gaps 1

QY 1014 cagtatcacaaaccttactcagttcttactacaactatctcaagaataaalygtgacat 1073

Db 2948 cactgcctcaatccaaccaccagattttagcagcgttcaagccgtatccagacgaact 3007

QY 1074 ggaacagtgtaactgactccctggtgcactcttgcagaatcaacttaactccctagcagcgt 1133

Db 3008 caacgaagtcgcaaaaatcaattaccacacttagaanaaagtaactgactcgtttgtctgtagt 3067

QY 1134 agtccctcaaatccgaagaacttttagactgttcaacgcgcgaagaagggggaactgttt 1193

Db 3068 agtccctcagaacccgaagagagcctgagattgtcctctcctaaagaagggaagcctctgc 3127

QY 1194 atttttagagagaagaagcgtcttatattgttaatacatccagaattgtcaactgagaagaat 1253

Db 3128 agccttaagaagaagatgtgttttatgtcagaacacacaggagactagtgagagagaagat 3187

QY 1254 taaagaagaatctcgagatctcgaaatacaatggtgagcagagagcctcaaaaaccggaagcgtg 1313

Db 3188 ggcacaaactaagggaaagagccttaatacagagacaaaactatttggatcagggccaaggttg 3247

QY 1314 gggcctcct-----cagccaatgagatgcccctggcttcctccctcttgg 1358

Db 3248 gttcgaagggcagtttatatgatacccccctggttaccaccttaactccacacatacaggg 3307

QY 1359 acctcagcagctcaatatattgttactcctcttggaccctgtatcttaaacctcctgtt 1418

Db 3308 acctcaatagtaactcttactgacttactactcttggaccctgcattccaatcagttagt 3367

QY 1419 taagtgtctctctccgaatttaagctgaaag 1452

Db 3368 tcaatttgttaagaagcaggtctcagtaattccag 3401

```

1 RESULT 12
2 US-09-011-745-5
3 Sequence 5, Application US/09011745
4 Patent No. 6165715
5 GENERAL INFORMATION:
6 APPLICANT: Collins, Mary KL
7 APPLICANT: Weils, Robin A
8 APPLICANT: Takeuchi, Yasuhiro
9 APPLICANT: Cosset, Francois-Ioic
10 TITLE OF INVENTION: Expression systems
11 FILE REFERENCE: 09/011,745
12 CURRENT APPLICATION NUMBER: US/09/011,745
13 CURRENT FILING DATE: 1998-06-22
14 EARLIER APPLICATION NUMBER: PCT/GB96/02061
15 EARLIER FILING DATE: 1996-08-23
16 EARLIER APPLICATION NUMBER: GB9517263.1
17 EARLIER FILING DATE: 1995-08-23
18 NUMBER OF SEQ ID NOS: 29
19 SOFTWARE: PatentIn Ver. 2.0
20 SEQ ID NO 5
21 LENGTH: 6028
22 TYPE: DNA
23 ORGANISM: Artificial Sequence
24 FEATURE:
25 OTHER INFORMATION: Description of Artificial Sequence: Portion of
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: (3774)
29 OTHER INFORMATION: n is any nucleotide
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: (3775)
33 OTHER INFORMATION: n is any nucleotide
34 FEATURE:
35 NAME/KEY: misc_feature
36 LOCATION: (3776)
37 OTHER INFORMATION: n is any nucleotide
38 FEATURE:
39 NAME/KEY: misc_feature
40 LOCATION: (3777)
41 OTHER INFORMATION: n is any nucleotide

```



```
ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 15280-128-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10970 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 1..10970
: OTHER INFORMATION: /standard_name="p537 retroviral
: OTHER INFORMATION: vector"
: US-08-716-351A-5

Query Match 4.68; Score 68; DB 3; Length 10970;
Best Local Similarity 56.18; Pred. No. 1.9e-11;
Matches 128; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1087 gaccctgtcaccctgcaagatcaactaactcccttagagcagatgctctcaaat 1146
Db 6659 GACTCATACAGCAAGCTAGAGGACTCAGCTCCCTATGTGAGTAGTACTCCAAAT 6918
QY 1147 cgaagctttagactgtcaccgccaagaagggggaacctgttatttttagagaa 1206
Db 6919 AGGAGAGGCGCTTGACTACTTCTTAAAGAGAGGCGCTCGCGCCCTTAAAGAA 6978
QY 1207 gaaagctgtattatgttaacatccaagaatgtcactgagaaagttaaagaattcga 1266
Db 6979 GAGGCTGTTTATGTAGACCACTCAGCTGACAGTACGAGACTCCATGAAAAAACTTMAA 7038
QY 1267 gatcaatacatgtagacagagagagcttcaaaacacccggaagcttg 1314
Db 7039 GAAAGACTAGTAAAGACAGCTTAGAGCCCGCCAGAAAAACCAAAATGG 7086

RESULT 15
US-08-258-420-8
: Sequence 8, Application US/08258420
: Patient No. 5710037
: GENERAL INFORMATION:
: APPLICANT: Nienhuis, Arthur W.
: APPLICANT: Vanin, Ello F.
: TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Carrella, Byrne, Baln, Gilfillan, Cecchi, Stewart & Olstein
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/258,420
: FILING DATE: 10-JUN-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Olstein, Elliot M.
```

```
REGISTRATION NUMBER: 24,025
: REFERENCE/DOCKET NUMBER: 271010-208
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1911 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: FEATURE:
: NAME/KEY: retroviral envelope sequence
: US-08-258-420-8

Query Match 4.58; Score 67; DB 1; Length 1911;
Best Local Similarity 48.68; Pred. No. 1.5e-11;
Matches 227; Conservative 0; Mismatches 225; Indels 15; Gaps 1;

QY 1030 actcagttctactcaactctccaagaataatgtgacatggaacaggtcactgac 1089
Db 1417 ACGCAGCAGTTTGACACCTTCATGCCGCTATCCAGACGACCTCAAGAGTCAAGAAAG 1476
QY 1090 tccctgtcaccctgcaagatcaactaactcccttagagcagatgctctcaaatga 1149
Db 1477 TCAATTACCAACCTAGAAAATCACTGACCTGTTGTGAGTAGTCTTACAGAACCGC 1536
QY 1150 agaacttagactgtcaccgccaagaagggggaacctgttatttttagagaaga 1209
Db 1537 AGAGGCCCTAGATTGCTATTCTTAAAGSAGGAGGTCTCTCGCAGAGCCCTTAAAGAGAA 1596
QY 1210 cgtgttattatgttaacatccaagaatgtcactgagaaagttaaagaattcagat 1269
Db 1597 TGTGTTTATATGACAGACACACGCGGCTAGTAGAGACACACATGCGCCAAATTAAGAAA 1656
QY 1270 cgaatacatgtagacagagagagcttcaaaacacccggaagcttg 1332
Db 1657 AGCCTTAATCAGAGCAAAAACCTATTGTGACAGAGCCCAAGAGATGTTGGAAGGCTGTTT 1716
QY 1323 -----cagccaatggatgccctgggtlctccctcttagagaccttagcagctcta 1374
Db 1717 AATGATCCCGCTTGTTTACCACCTTAATCTCCACATCATGAGGACCTTAATAGTACTC 1776
QY 1375 atattgtacctcctcttggagccctgtatctttaacctccttgaagttgtctctcc 1434
Db 1777 TTACTGATCTTACTCTTGGACCTTGCAATCTCAATCGATTAGTCAATTGTTAAAGAC 1836
QY 1435 agaattgagctgtaaagctacagatggtcttaacaatgagaccoca 1481
Db 1837 AGGATATCAGTGTGTCAGGCTGTGATTGACTCAACAATATTCACCA 1883
```

Search completed: June 20, 2002, 09:38:01
Job time: 8391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:42:19 ; Search time 443.18 Seconds

(without alignments)
5737.511 Million cell updates/sec

Title: US-09-319-156a-9

Perfect score: 1481

Sequence: 1 atgcccctcctatcatcac.....gtctacaagaatgaaccacca 1481

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Optimal number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_032802.*
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1481	100.0	1481	19	AAV43217
2	1481	100.0	1481	20	AAV43217
3	1481	100.0	1481	21	AAV43217
4	1461.8	98.7	2030	21	AAV43217
5	1335.4	90.2	2946	20	AAV43217
6	1335.4	90.2	2946	21	AAV43217
7	1333.8	90.1	1617	22	AAH20070
8	1330.6	89.8	2781	22	AAH20070
9	1324.2	89.4	7582	21	AAH20070

10	1322.6	89.3	7582	20	AAH20069
11	1308.2	88.3	2782	22	AAH20069
12	1306.6	88.2	2782	22	AAH20069
13	1306.6	88.2	2782	21	AAH20069
14	1311.6	76.4	1894	22	ABH45822
15	1311.6	76.4	1894	22	ABH45822
16	1311.6	76.4	1894	22	ABH45822
17	1311.6	76.4	1894	22	ABH45822
18	1311.6	76.4	1894	22	ABH45822
19	1311.6	76.4	1894	22	ABH45822
20	1311.6	76.4	1894	22	ABH45822
21	1311.6	76.4	1894	22	ABH45822
22	1311.6	76.4	1894	22	ABH45822
23	1311.6	76.4	1894	22	ABH45822
24	1311.6	76.4	1894	22	ABH45822
25	1311.6	76.4	1894	22	ABH45822
26	1311.6	76.4	1894	22	ABH45822
27	1311.6	76.4	1894	22	ABH45822
28	1311.6	76.4	1894	22	ABH45822
29	1311.6	76.4	1894	22	ABH45822
30	1311.6	76.4	1894	22	ABH45822
31	1311.6	76.4	1894	22	ABH45822
32	1311.6	76.4	1894	22	ABH45822
33	1311.6	76.4	1894	22	ABH45822
34	1311.6	76.4	1894	22	ABH45822
35	1311.6	76.4	1894	22	ABH45822
36	1311.6	76.4	1894	22	ABH45822
37	1311.6	76.4	1894	22	ABH45822
38	1311.6	76.4	1894	22	ABH45822
39	1311.6	76.4	1894	22	ABH45822
40	1311.6	76.4	1894	22	ABH45822
41	1311.6	76.4	1894	22	ABH45822
42	1311.6	76.4	1894	22	ABH45822
43	1311.6	76.4	1894	22	ABH45822
44	1311.6	76.4	1894	22	ABH45822
45	1311.6	76.4	1894	22	ABH45822

ALIGNMENTS

RESULT 1	
AAV43217	
AAV43217 standard; CDNA: 1481 BP.	
AC AAV43217;	
DT 29-DEC-1998 (first entry)	
XX Multiple sclerosis associated retrovirus fragment 5.	
DE Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;	
KW gag gene; env gene; Rheumatoid arthritis-associated virus; ss.	
KM Multiple sclerosis associated retrovirus.	
OS Multiple sclerosis associated retrovirus.	
XX Key	Location/Qualifiers
FT CDS	1..1479
FT	/*tag= a
FT	/product= "Encodes protein AAV71068"
FT	/transl_except= (pos:115-117, appears to code for a
FT	stop codon)
FT	/note= "CDS does not contain a stop codon"
XX W09823755-A1.	
XX 04-JUN-1998.	
XX 26-NOV-1997; 97WO-IB01482.	
XX 26-NOV-1996; 96US-0756429.	
XX (INMR) BIO MERIEUX.	

Complete human end
HERV-W envelope pr
Human endogenous r
5' non coding, 3'
Human breast cell
Human foetal liver
Probe #4444 for ge
Human brain expres
Human bone marrow
Probe #4541 for ge
Probe #4666 used t
Probe #4413 used t
Human endogenous r
5' non coding, 3'
DNA encoding novel
Human breast cell
Human foetal liver
Probe #14338 for g
Human brain expres
Human bone marrow
Probe #13736 for g
Probe #17801 used
Probe #9403 used t
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Human endogenous r
Probe Penv-C15 use
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Human endogenous r
Human endogenous r

XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
PI Mandrand B, Paranhos-Baccala G, Perron H;
XX WPI: 1998-322732/28.
DR P-PSDB; AAW71068.
XX
XX
PT New nucleic acid from retroviruses - useful for diagnosis,
PT prevention and treatment of, e.g. multiple sclerosis
XX
PS Disclosure; Pages 184-185; 286pp; English.
XX
XX
CC The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) genomic fragment used in the method of the
CC invention. The invention provides complete or partial genomic
CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC polypeptides encoded by these genes. The invention also provides
CC antibodies raised against the polypeptides. The genomic sequences,
CC polypeptides and antibodies are also claimed useful for diagnosing
XX infection by MS and rheumatoid arthritis-associated viruses, and also
XX for prevention and treatment of infection with these viruses.
XX
XX Sequence 1481 BP; 412 A; 410 C; 261 G; 398 T; 0 other;

Qy	661	ggacctcttgtttccaatctcggaaataaacccatacctcaaacctcctgtgtataatt	720
Db	661	ggacctcttgtttccaatctcggaaataaacccatacctcaaacctcctgtgtataatt	720
Qy	721	agcaatactatagaacaacaacagctcccaatgcatcagtggtgtaaacctcccaaga	780
Db	721	agcaatactatagaacaacaacagctcccaatgcatcagtggtgtaaacctcccaaga	780
Qy	781	atagctcgcgaacccctcagaataattttttctcgtgtaacctcagaactaaattttg	840
Db	781	atagctcgcgaacccctcagaataattttttctcgtgtaacctcagaactaaattttg	840
Qy	841	aatggcctctcagaatactatgtctcctcctcatctcttagtgccctatgacaactac	900
Db	841	aatggcctctcagaatactatgtctcctcctcatctcttagtgccctatgacaactac	900
Qy	901	actgaacaagaattatatacatcatcgtctgaactaagcccaacaacaagaattaccatt	960
Db	901	actgaacaagaattatatacatcatcgtctgaactaagcccaacaacaagaattaccatt	960
Qy	961	cttcctcttctgtatcagaagcagagctcgaagcagactagtgatctggtcagatctc	1020
Db	961	cttcctcttctgtatcagaagcagagctcgaagcagactagtgatctggtcagatctc	1020

Query Match	100.0%	Score 1481;	DB 19;	Length 1481;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1481; Conservative	0;	Mismatches	0;	Gaps 0

Db 1021 acaactctactcagttctactacaactatctccaagaataatgtygcactggaacag 1080

QY	1	atgagccccccttataaactttctcttactctgttctcttaaccccttcctcctact	60
Db	1	atgagccccccttataaactttctcttactctgttctcttaaccccttcctcctact	60
QY	61	gaaccccccctcaatgcgtgtgtacaaacagatgctccctctaccaaagattcttatgaga	120
Db	61	gaaccccccctcaatgcgtgtgtacaaacagatgctccctctaccaaagattcttatgaga	120
QY	121	acgagcgtctcccttgaaatatgtatgagccccaatataatgaggtttatatgaagaaactcc	180
Db	121	acgagcgtctcccttgaaatatgtatgagccccaatataatgaggtttatatgaagaaactcc	180
QY	181	acctcaactctgcacccaacccaatctgcccgcgaactctgtatactctgcacactcttgatg	240
Db	181	acctcaactctgcacccaacccaatctgcccgcgaactctgtatactctgcacactcttgatg	240
QY	241	catctcaaatatctcttatgttgagacgggaaaaatgttatactcgtgtctcttgagagact	300
Db	241	catctcaaatatctcttatgttgagacgggaaaaatgttatactcgtgtctcttgagagact	300
QY	301	gagagccaaactctgtctgtgacttacttcaacccaataccagatctgtctgaatggggatg	360
Db	301	gagagccaaactctgtctgtgacttacttcaacccaataccagatctgtctgaatggggatg	360
QY	361	caaaagtctagagaaagaaaaacaagtataaggaagaacatctcccaactgacccggggagcat	420
Db	361	caaaagtctagagaaagaaaaacaagtataaggaagaacatctcccaactgacccggggagcat	420
QY	421	agcaccctctagccctctcaaaagagactagtctctctcaaaactatcatgaaacctctgtaac	480
Db	421	agcaccctctagccctctcaaaagagactagtctctctcaaaactatcatgaaacctctgtaac	480
QY	481	catatctgcgcctgggagcctattataatacaacccctctgctgcctcatgaggtcttaagcc	540
Db	481	catatctgcgcctgggagcctattataatacaacccctctgctgcctcatgaggtcttaagcc	540
QY	541	caaaaacccttaactctgtctgaatgtgctctcccccctgtgactctcagccatcatcttcaatc	600
Db	541	caaaaacccttaactctgtctgaatgtgctctcccccctgtgactctcagccatcatcttcaatc	600
QY	601	cccttctcttgaaacaaatggaaaactcttagcacagaaataaaccacactcttcggtttagtta	660
Db	601	cccttctcttgaaacaaatggaaaactcttagcacagaaataaaccacactcttcggtttagtta	660

QY	1081	gtcaactgaccccttgctcaactcttgcaagaatcaacttaactcccttaagcagagtagtcctt	1140
Db	1081	gtcaactgaccccttgctcaactcttgcaagaatcaacttaactcccttaagcagagtagtcctt	1140
QY	1141	caaaaatcgaagaagctttagaactctgctcaacccgcaaaaagaggggaaacctgttatcttta	1200
Db	1141	caaaaatcgaagaagctttagaactctgctcaacccgcaaaaagaggggaaacctgttatcttta	1200
QY	1201	ggaggaagaagcgcgtctattatgtttaatccaatccaagatgtctaccttagaagaagttaagaa	1260
Db	1201	ggaggaagaagcgcgtctattatgtttaatccaatccaagatgtctaccttagaagaagttaagaa	1260
QY	1261	atctcgagatcgaataacatgcttagagagagaggaagcttcaaaaacccgaaacgtctggggcttc	1320
Db	1261	atctcgagatcgaataacatgcttagagagaggaagcttcaaaaacccgaaacgtctggggcttc	1320
QY	1321	ctcagcgaatctgtagtgcgccggggtctcccccctctctttagagaccttaagcagctctaatactg	1380
Db	1321	ctcagcgaatctgtagtgcgccggggtctcccccctctctttagagaccttaagcagctctaatactg	1380
QY	1381	ttaactctctcttggaaccccgctgatactcttaaacctctcctgttaagcttctctctccagaat	1440
Db	1381	ttaactctctcttggaaccccgctgatactcttaaacctctcctgttaagcttctctctccagaat	1440
QY	1441	gaagcctgataagctacagatgctgcttacaagaatggaacccca	1481
Db	1441	gaagcctgataagctacagatgctgcttacaagaatggaacccca	1481
RESULT 2			
AAK29703			
ID	AAK29703 standard; DNA; 1481 BP.		
XX			
AC	AAK29703;		
XX			
DT	08-JUN-1999 (first entry)		
XX			
DE	Clone C15 from MSRV-1.		
XX			
KM	Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;		
XX	rheumatoid polyarthritis; ss.		
OS	Multiple sclerosis related virus type 1.		
XX			
PN	FR2765588-A1.		

XX	08-JAN-1999.
PD	07-JUL-1997; 97FR-0008816.
XX	
PF	07-JUL-1997; 97FR-0008816.
XX	
PR	07-JUL-1997; 97FR-0008816.
XX	
PA	(INMR) BIO MERIEUX.
XX	
DR	WPI; 1999-098275/09.
DR	P-PsDB; AAM9553.
XX	
PT	Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT	multiple sclerosis or rheumatoid polyarthritis
XX	
PS	Claim 1; Page 37-38; 83pp; French.
XX	
CC	This sequence represents clone C15 from a novel multiple sclerosis
CC	related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC	prophylactic or therapeutic compositions to inhibit expression of a
CC	multiple sclerosis related virus and/or virus associated with
CC	rheumatoid polyarthritis.
XX	
SQ	Sequence 1481 BP: 412 A; 410 C; 261 G; 398 T; 0 other:
Query Match	100.0%; Score 1481; DB 20; Length 1481;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1481; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 atggccctccttataactttctcttactgttcttacccttcgctcatc 60
Db	1 atggccctccttataactttctcttactgttcttacccttcgctcatc 60
OY	61 gacccccctcatagtctgtgtacaaccagtagctcccttaccagaagtltatgaaga 120
Db	61 gacccccctcatagtctgtgtacaaccagtagctcccttaccagaagtltatgaaga 120
OY	121 acgcggtctcccggaatatgatgcccatcataatagaatttatcaaggaaacctc 180
Db	121 acgcggtctcccggaatatgatgcccatcataatagaagtltatcaaggaaacctc 180
OY	181 accttaactgcccacaccatalatgcccgcgaactgctataactctgtcscactcttgatg 240
Db	181 accttaactgcccacaccatalatgcccgcgaactgctataactctgtcscactcttgatg 240
OY	241 catgcaaatatactatatttgacagaggaataatgatataactcagtgtgtccgagagact 300
Db	241 catgcaaatatactatatttgacagaggaataatgatataactcagtgtgtccgagagact 300
OY	301 ggagccaactgttgttgactactactaaccccatcacatgtctcgaaggagggtgaatt 360
Db	301 ggagccaactgttgttgactactactaaccccatcacatgtctcgaaggagggtgaatt 360
OY	361 caaagtcagcgcaagaagaanaacaaglaaaggagaactccccaaactgaccccgaggacat 420
Db	361 caaagtcagcgcaagaagaanaacaaglaaaggagaactccccaaactgaccccgaggacat 420
OY	421 agcacccctagagccctacaagaagactagtctctcaaaaactacagaaacctccgctac 480
Db	421 agcacccctagagccctacaagaagactagtctctcaaaaactacagaaacctccgctac 480
OY	481 cattaactgcctgtgtagcctattttaaccacccctcaactggctccatagaggtccagcc 540
Db	481 cattaactgcctgtgtagcctattttaaccacccctcaactggctccatagaggtccagcc 540
OY	541 caaaacctataactgtgtgagatgycctccccctgcaactltaaggcatalacttcaatc 600
Db	541 caaaacctataactgtgtgagatgycctccccctgcaactltaaggcatalacttcaatc 600
OY	601 cctgttctctgaacaatggaacaacttcagcacagaataaacaacacttcgltttagta 660
Db	601 cctgttctctgaacaatggaacaacttcagcacagaataaacaacacttcgltttagta 660

QY	661	ggacctcttggttccaatctcggaaataaccctaacctcaactcaactcgtgtgtaaat	720
Db	661	ggacctcttggttccaatctcggaaataaccctaacctcaactcaactcgtgtgtaaat	720
QY	721	agcaatactatagacaacaacagctccccaatgcatacagttggttaacactccacaga	780
Db	721	agcaatactatagacaacaacagctccccaatgcatacagttggttaacactccacaga	780
QY	781	atagctctgctaaccctcaggaaataattttgtctgtgtgtaaccagctcatatgtttg	840
Db	781	atagctctgctaaccctcaggaaataattttgtctgtgtgtaaccagctcatatgtttg	840
QY	841	aatgctcttcaagaatactatgtgtcttccctcaattcttagtgcacctatgacatcac	900
Db	841	aatgctcttcaagaatactatgtgtcttccctcaattcttagtgcacctatgacatcac	900
QY	901	actgacaagaattatatacatcatgtgtatcctcaagccccacaacaaagagtaacctt	960
Db	901	actgacaagaattatatacatcatgtgtatcctcaagccccacaacaaagagtaacctt	960
QY	961	cttcctcttggtatcagaagcagaggtgtctaggcagactagtaactgtgcattgycag	1020
Db	961	cttcctcttggtatcagaagcagaggtgtctaggcagactagtaactgtgcattgycag	1020
QY	1021	acaacctactactgcttctactacaactctcccaagaataatggtgtacatggaacag	1080
Db	1021	acaacctactactgcttctactacaactctcccaagaataatggtgtacatggaacag	1080
QY	1081	gtcactgtactccctgtgcacctcttgcaagatcaacttaactccctagcagcagtagctt	1140
Db	1081	gtcactgtactccctgtgcacctcttgcaagatcaacttaactccctagcagcagtagctt	1140
QY	1141	caaaatcgaaagcgtcttagactgtgtcaaccgccaagaagggggaaacctgttatttta	1200
Db	1141	caaaatcgaaagcgtcttagactgtgtcaaccgccaagaagggggaaacctgttatttta	1200
QY	1201	ggagaagaagcgtcttattatgttaataccaatgtcaactgcagagaagttaaga	1260
Db	1201	ggagaagaagcgtcttattatgttaataccaatgtcaactgcagagaagttaaga	1260
QY	1261	attcagatcgatacaatgttagacgaggaagcttcaaaacacggaaagctcggggctc	1320
Db	1261	attcagatcgatacaatgttagacgaggaagcttcaaaacacggaaagctcggggctc	1320
QY	1321	ctcagccaatgtagtccctgtgtctccctcttagagccctcagcagctcataattg	1380
Db	1321	ctcagccaatgtagtccctgtgtctccctcttagagccctcagcagctcataattg	1380
QY	1381	ttaccctcctttggaacctgtatctttaacctctctgttaagttgtctcttcagaatt	1440
Db	1381	ttaccctcctttggaacctgtatctttaacctctctgttaagttgtctcttcagaatt	1440
QY	1441	gaagctgtaagctacagatcgtgtcttaacaatggaaccca	1481
Db	1441	gaagctgtaagctacagatcgtgtcttaacaatggaaccca	1481
RESULT 3			
AAA63826			
ID AAA63826 standard; DNA; 2030 BP.			
XX			
AC AAA63826;			
XX			
XX 04-DEC-2000 (first entry)			
XX			
DE Nucleotide sequence of the MSRV-1 3' env and LTR regions.			
XX			
KM MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;			
XX ss.			
OS Multiple Sclerosis retrovirus 1.			
XX			

```

FH Key Location/Qualifiers
FT CDS 1..1629
FT sig_peptide /note= "Contains one termination codon"
FT 1..81
FT CAAAT_signal /tag= b
FT 1800..1807
FT CAAAT_signal /tag= c
FT 1858..1864
FT TATA_signal /tag= d
FT 1906..1911
FT polyA_signal /tag= e
FT 1996..2002
FT /tag= f

WO200047745-A1.
17-AUG-2000.
15-FEB-2000; 2000WO-1B00159.
15-FEB-1999; 99EP-0420041.
(INMR ) BIO MERIEUX.
Paranhos-Baccala G, Perron H, Komurian-Pradel F;
WPI; 2000-506097/45.
P-PSDB; AAB08195.

Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
biological sample
PS Disclosure: Fig 2; 23pp; English.

The present sequence represents the nucleotide sequence corresponding
to the 3' env region and long terminal repeat sequences from clone
CL6 of Multiple Sclerosis retrovirus (MSRV-1). The specification
describes a long terminal repeat (LTR)-RUS region which encodes the
expression of a MSRV-1 protein. This is unusual for LTRs, in
particular in the RUS region. The sequence includes CAAAT and TATA
signals which are present in the U3 and R regions and are not directed
towards the CDS indicated in the features table. Probes and antibodies
to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
are used to detect the presence of MSRV-1 retrovirus in a biological
sample.

Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 other:

Query Match 100.0%; Score 1481; DB 21; Length 2030;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 241 catgcaaatatcatattatggacagggaatgatattccatgtgtctcggagactt 300
Qy 301 ggaagcaactgtctgtgagacttacttcaaccataccagatgtctgattgagggtgaatt 360
Db 301 ggaagcaactgtctgtgagacttacttcaaccataccagatgtctgattgagggtgaatt 360
Qy 361 caaggtcagcagaagagaanaaacagtaagaagcaattcccaactgacccgggagac 420
Db 361 caaggtcagcagaagagaanaaacagtaagaagcaattcccaactgacccgggagac 420
Qy 421 agcaccctcagccctcagaagaagactagttctcctaaactcctatgaacccctctacc 480
Db 421 agcaccctcagccctcagaagaagactagttctcctaaactcctatgaacccctctacc 480
Qy 481 catctcgcctgtgtagccatttaatacacccctcactcgcctcatgaggtctcagcc 540
Db 481 catctcgcctgtgtagccatttaatacacccctcactcgcctcatgaggtctcagcc 540
Qy 541 caaaacccctactaactgttggatgtgctccctcgcctcactgaagccatataatc 600
Db 541 caaaacccctactaactgttggatgtgctccctcgcctcactgaagccatataatc 600
Qy 601 cctgttccctggaacaatggacaacttcagacagagaatgaacacactccgttttag 660
Db 601 cctgttccctggaacaatggacaacttcagacagagaatgaacacactccgttttag 660
Qy 661 ggaacctctgttcccaatcctggaataacccataccctcaacctcactgtttaaat 720
Db 661 ggaacctctgttcccaatcctggaataacccataccctcaacctcactgtttaaat 720
Qy 721 agcaatactatagaacaacacagctcccaatgcatcaggttgggttaacacctccaca 780
Db 721 agcaatactatagaacaacacagctcccaatgcatcaggttgggttaacacctccaca 780
Qy 781 atagctcgtcctccctcagaagaatttttctgtcgtgagactcagccatattgtttg 840
Db 781 atagctcgtcctccctcagaagaatttttctgtcgtgagactcagccatattgtttg 840
Qy 841 aatgctctcctcagaactctatgtgctcctcctatctcattgtgccccctatgacatc 900
Db 841 aatgctctcctcagaactctatgtgctcctcctatctcattgtgccccctatgacatc 900
Qy 901 actgacaagaattatatacaatcatatgtctgaactaagccccacaacaaagattaccat 960
Db 901 actgacaagaattatatacaatcatatgtctgaactaagccccacaacaaagattaccat 960
Qy 961 cttcttctgtatcagaacagagatgtcaggaactgaactgaactgacatgacatgac 1020
Db 961 cttcttctgtatcagaacagagatgtcaggaactgaactgaactgacatgacatgac 1020
Qy 1021 acaacctactcagttctactacataactatctcaagaataaatggtgacatggacag 1080
Db 1021 acaacctactcagttctactacataactatctcaagaataaatggtgacatggacag 1080
Qy 1081 gtcactgactccctgtgctccttgcacagatcaacttaactccctcagcagatgctct 1140
Db 1081 gtcactgactccctgtgctccttgcacagatcaacttaactccctcagcagatgctct 1140
Qy 1141 caaaatcgaagaagctttaactgtcgaaccgcaaaaggggggaacccgtttatttta 1200
Db 1141 caaaatcgaagaagctttaactgtcgaaccgcaaaaggggggaacccgtttatttta 1200
Qy 1201 ggaagaagaagctgttatatgtttaatcaatccagaatgtgtaacgagagaagttaaga 1260
Db 1201 ggaagaagaagctgttatatgtttaatcaatccagaatgtgtaacgagagaagttaaga 1260
Qy 1261 attcgaatcgaatacaaatgtgtagcagaaggagcttcaaaacgaaacgctgaggcc 1320
Db 1261 attcgaatcgaatacaaatgtgtagcagaaggagcttcaaaacgaaacgctgaggcc 1320
Qy 1321 ctcaagcaatgtagtgcctgggttctccctctctagagcctctagcagcttaatatg 1380
Db 1321 ctcaagcaatgtagtgcctgggttctccctctctagagcctctagcagcttaatatg 1380

```


QY 1381 ttactcctttgaccctgatatcttaacctgtttaaagttgtctctccagaatt 1440
|||||
Db 1381 ttactcctttgaccctgatatcttaacctgtttaaagttgtctctccagaatt 1440
QY 1441 gaagctgtaagctacagatggtcttaacaaatggaaccca 1481
|||||
Db 1441 gaagctgtaagctacagatggtcttaacaaatggaaccca 1481
RESULT 4
AAA96625 standard; DNA: 1629 BP.
XX
XX AAA96625:
AC
XX 08-FEB-2001 (first entry)
DT
XX
DE DNA encoding an envelope (env) protein of MSRV-1.
XX
XX Envelope protein: MSRV-1; superantigen: autoimmune disease; Vbeta16;
XX Vbeta17; multiple sclerosis; vaccine; ss.
XX
XX Human endogenous retrovirus.
OS
XX
XX Location/Qualifiers
FH Key 1.1629
FT CDS /*tag= a
FT /product= "envelope protein"
XX
XX MO200057185-A1.
PD 28-SEP-2000.
XX
XX 20-MAR-2000; 2000MO-FR00691.
PF
XX
XX 19-MAR-1999; 99FR-0003622.
PR 28-OCT-1999; 99FR-0013755.
XX
XX
PA (INMR) BIO MERIEUX.
XX
XX Perron H, Lafont M;
PI
XX
XX WPI: 2000-638279/61.
DR P-PSDB; AAB19069.
XX
PT Detecting superantigen activity, useful for identifying agents for
PT treatment or prevention of autoimmune disease, from expansion or loss
PT of particular lymphocyte Vbeta determinants -
XX
PS Claim 25; Page 123-124; 134pp; French.
XX
XX The present sequence encodes an envelope protein of MSRV-1. The
XX envelope protein expressed by the endogenous human retrovirus MSRV-1
XX has superantigen activity associated with autoimmune disease. The
XX protein can be detected using the method of the invention. The
XX specification describes a process for detecting activity of a
XX superantigen in a biological sample. The process comprises identifying
XX large scale expansion or loss of lymphocytes that carry at least one
XX of the Vbeta16 and/or Vbeta17 determinants. The method is used to
XX screen for agents that inhibit the superantigen, especially those
XX associated with MSRV-1 which is implicated in autoimmune disease,
XX particularly multiple sclerosis. These agents are potentially useful
XX for treatment or prevention (e.g. as vaccines) of autoimmune diseases.
XX
SQ Sequence 1629 BP; 456 A; 454 C; 287 G; 432 T; 0 other;

Query Match 98.7%; Score 1461.8; DB 21; Length 1629;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 atggccctcctatcatattcttcttactgttctcttaacccttgcgtctact 60

Db 1 atggccctcctatcatattcttcttactgttctcttaacccttgcgtctact 60
QY 61 gcacccctcctatctgtctgtacaacagtagtcccttccacaagattctatgaaga 120
Db 61 gcacccctcctatctgtctgtacaacagtagtcccttccacaagattctatgaaga 120
QY 121 acgagctcttcgtgaaatgatgagcccatcatatagagattatctcaagaggaactcc 180
Db 121 acgagctcttcgtgaaatgatgagcccatcatatagagattatctcaagaggaactcc 180
QY 181 acccttaccgcccacaccatattgcccgcgaactgtataactctgccaactcttgatg 240
Db 181 acccttaccgcccacaccatattgcccgcgaactgtataactctgccaactcttgatg 240
QY 241 catgcaaatcatcatattatgacaggggaaatgatataactctagtgtctctgagagact 300
Db 241 catgcaaatcatcatattatgacaggggaaatgatataactctagtgtctctgagagact 300
QY 301 ggaagccactgtctgttggacttacttcaaccatccagtagtctgtatgggggtggaatt 360
Db 301 ggaagccactgtctgttggacttacttcaaccatccagtagtctgtatgggggtggaatt 360
QY 361 caaggtcagcagaagaaacaagtaagaagcaatctcccaactgacccgggagacat 420
Db 361 caaggtcagcagaagaaacaagtaagaagcaatctcccaactgacccgggagacat 420
QY 421 agcacccctagcccttacaagaagactgtctctcaaaactacatgaaaacctccgtacc 480
Db 421 agcacccctagcccttacaagaagactgtctctcaaaactacatgaaaacctccgtacc 480
QY 481 catatcgccctggtgagacttattatatacccttactctggtccatagaagttctcagcc 540
Db 481 catatcgccctggtgagacttattatatacccttactctggtccatagaagttctcagcc 540
QY 541 caaaaccctactaactgttggatgtgctccctccctgcactcagggccatacttccaatc 600
Db 541 caaaaccctactaactgttggatgtgctccctccctgcactcagggccatacttccaatc 600
QY 601 cctgttctgaaacaatggacaacttccagcagaagaataaacaacactccgttttagta 660
Db 601 cctgttctgaaacaatggacaacttccagcagaagaataaacaacactccgttttagta 660
QY 661 ggaactctgttccaatcttgaaataaccatccctcaaacctccactcgtgtataattt 720
Db 661 ggaactctgttccaatcttgaaataaccatccctcaaacctccactcgtgtataattt 720
QY 721 agcaatactatagacacacacagctcccaatgcatcaggtgtgtataacccctccacaga 780
Db 721 agcaatactatagacacacacagctcccaatgcatcaggtgtgtataacccctccacaga 780
QY 781 atagtctgcttaacctcagaatatatttctgtgtgtgtactcgaactcatattgtttg 840
Db 781 atagtctgcttaacctcagaatatatttctgtgtgtgtgtactcgaactcatattgtttg 840
QY 841 aatggctttagaatcatgtgcttctctcatattcttaagccccctatataccttac 900
Db 841 aatggctttagaatcatgtgcttctctcatattcttaagccccctatataccttac 900
QY 901 actgaacaagattatatacaatcatgtcttacttaagcccccaacaagaagtagtaccat 960
Db 901 actgaacaagattatatacaatcatgtcttacttaagcccccaacaagaagtagtaccat 960
QY 961 ctctcctttgttatcagagcagagtgctagggcagactagtagtactgtgcagatc 1020
Db 961 ctctcctttgttatcagagcagagtgctagggcagactagtagtactgtgcagatc 1020
QY 1021 acaaccttactcagttcttactatacaactatctcaagaataaataatgtgtgaatgtgaag 1080
Db 1021 acaaccttactcagttcttactatacaactatctcaagaataaataatgtgtgaatgtgaag 1080
QY 1081 gtcaactgactccctgtgtacacttgcagaatcaacttaactcctcctagcagtagtctct 1140
Db 1081 gtcaactgactccctgtgtacacttgcagaatcaacttaactcctcctagcagtagtctct 1140

QY 841 aatggtcttcagaatcatatgtcttctctcattcttagtgcctccatgaccatctac 900
 Db 1768 aatggtcttcagaatcatatgtcttctctcattcttagtgcctccatgaccatctac 1827
 QY 901 actggaataaattatatacatatgtctgtaactaagccccaacaagaatgaccatt 960
 Db 1828 actggaataaattatatacatatgtctgtaactaagccccaacaagaatgaccatt 1887
 QY 961 ctctcttcttctacagagagagtgtaggcagactagtagtactgtagtgcagatc 1020
 Db 1888 ctctcttcttctacagagagagtgtaggcagactagtagtactgtagtgcagatc 1947
 QY 1021 acaactctactcagttctactatacaactatctcaagaataaattggtgacatggaacag 1080
 Db 1948 acaactctactcagttctactatacaactatctcaagaataaattggtgacatggaacag 2007
 QY 1081 gtcaactactcctgtgtcactctgcaagatcaactaactcctcctagcagagtagtccctt 1140
 Db 2008 gtctgcagactcctgtgtcactctgcaagatcaactaactcctcctagcagagtagtccctt 2067
 QY 1141 caaatcgaagagctttagaactgtctacacgccaagaagagggaggaactgtttatttta 1200
 Db 2068 caaatcgaagagctttagaactgtctacacgccaagaagagggaggaactgtttatttta 2127
 QY 1201 ggaagaagaacgctgttatttattatcaatccagaattgtcactggaagaattaaagaa 1260
 Db 2128 ggaagaagaacgctgttatttattatcaatccagaattgtcactggaagaattaaagaa 2187
 QY 1261 attcgaatcgaatacaaatgtagagcagagagagcttcaaaaacacgaacgctggagcctc 1320
 Db 2188 attcgaatcgaatacaaatgtagagcagagagagcttcaaaaacacgaacgctggagcctc 2247
 QY 1321 ctcaagcaatgagatgacctgtgtctccctctcttagaactctagacactgaacactaatttg 1380
 Db 2248 ctcaagcaatgagatgacctgtgtctccctctcttagaactctagacactgaacactaatttg 2307
 QY 1381 ttactcctcttggaacctgtatcttcaactcctctgttaagtgtgtctcttcacagaatt 1440
 Db 2308 ttactcctcttggaacctgtatcttcaactcctctgttaagtgtgtctcttcacagaatt 2367
 QY 1441 gaagctgttaagctacagatggtcttacaataggaaccca 1481
 Db 2368 gaagctgttaagctacagatggtcttacaataggaaccca 2408

RESULT 6
 AA259468
 AA259468 standard; cDNA; 2946 BP.
 AA259468;
 11-APR-2000 (first entry)
 Human secreted protein AJ172.2 polynucleotide sequence.
 Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
 placental pathology; metastasis inhibition; nutritional activity;
 immune stimulator; haematopoiesis regulator; tissue growth;
 tumour inhibitor; anti-inflammatory; clone AJ172.2; ATCC_96115;
 gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO960020-A1.
 XX
 PD 25-NOV-1999.
 XX
 PF 17-MAY-1999; 99MO-US10915.
 XX
 PR 18-MAY-1998; 98US-0080478.
 XX 20-OCT-1998; 98US-0175928.

PA (GEM) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Mi S, Treacy M;
 XX
 DR WPI: 2000-116311/10.
 DR P-PSDB: AAY67313.
 XX
 PT New polynucleotides encoding secreted cDNA libraries, used to develop
 products for the diagnosis and treatment of neoplastic disease
 XX
 PS Claim 14; Page 107-108; 149pp; English.
 XX
 CC This is the human secreted protein AJ172.2 nucleotide sequence, obtained
 CC from a human adult testes cDNA library. The invention relates to secreted
 CC human and murine proteins. The polynucleotides and proteins are predicted
 CC to have biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals. Detection of the levels of the proteins can be used for the
 CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
 CC which modulate the expression or function of the proteins may be used for
 CC treating a neoplastic disease and inhibiting metastasis. Other suggested
 CC activities include nutritional activity (e.g. in feeds), cytokine and
 CC cell proliferation/differentiation activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The
 CC polynucleotide sequences are also stated to be useful for gene therapy.
 XX
 SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
 Query Match 90.2%; Score 1335.4; DB 21; Length 2946;
 Best Local Similarity 93.9%; Pred. No. 0;
 Matches 1390; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 1 atggccctcccttactactcttcttactcttcttcttacccttcgctcact 60
 Db 928 atggccctcccttactactcttcttactcttcttcttacccttcgctcact 987
 QY 61 gcaacccctccatgctgtgtacacacagtagccctcccttaccagaagtctctatgaaga 120
 Db 988 gcaacccctccatgctgtgtacacacagtagccctcccttaccagaagtctctatgaaga 1047
 QY 121 acgcgctctcctggaatatgtatgcccacatataaggaggttacttaaggaactcc 180
 Db 1048 atgcagcgtcccggaatatgtatgcccacatataaggaggttacttaaggaactcc 1107
 QY 181 accttcactgcccacacacacatctcccggaactgcctataactctgcacctttgcatg 240
 Db 1108 accttcactgcccacacacacatctcccggaactgcctataactctgcacctttgcatg 1167
 QY 241 catgcaaatcatctatttggacaggggaaatgtaataccttagtctcctgaggaact 300
 Db 1168 catgcaaatcatctatttggacaggggaaatgtaataccttagtctcctgaggaact 1227
 QY 301 ggaacacatgctgttggacttacttcaaccatacagtagtctgtaggggtggaatt 360
 Db 1228 ggaacacatgctgttggacttacttcaaccatacagtagtctgtaggggtggaatt 1287
 QY 361 caaggtcagcgaagaagaacaagtaagaagaatctcccaactgacccggggagcat 420
 Db 1288 caaggtcagcgaagaagaacaagtaagaagaatctcccaactgacccggggagcat 1347
 QY 421 agcacccctagcccttacaagaagtagtctcctcaaaactatcatgaaacccctcgtacc 480
 Db 1348 agcacccctagcccttacaagaagtagtctcctcaaaactatcatgaaacccctcgtacc 1407
 QY 481 catatcgtcctgtgtgagctatttaatacaccctcactcgtgctcatgaggtcagcc 540
 Db 1408 catatcgtcctgtgtgagctatttaatacaccctcactcgtgctcatgaggtcagcc 1467

Oy 541 caaaacccctactactgtgtgagatgtgctcccccctgcaacttcaaggccatatacttcaatc 600
|||||
Db 1468 caaaacccctactactgtgtgagatgtgctcccccctgcaacttcaaggccatatacttcaatc 1527
Oy 601 cctgtctccctggaacatgtggaacacttcagcagagaataaacaacactccgttttagta 660
|||||
Db 1528 cctgttacttggaacatgtggaacacttcagcagagaataaacaacactccgttttagta 1587
Oy 661 ggaactctgttctccaatctggaataaaccataccctcaaacctcactgtgttaaat 720
|||||
Db 1588 ggaactctgttctccaatctggaataaaccataccctcaaacctcactgtgttaaat 1647
Oy 721 agcaatatactatagacaacaacagctcccaatgcatcaggtgggttaacactcccaaga 780
|||||
Db 1648 agcaatatactatagacaacaacagctcccaatgcatcaggtgggttaacactcccaaga 1707
Oy 781 ataagctctgctccctcctcaggaatatttttctgtctgtgtactcagcctatctgtctg 840
1708 ataagctctgctccctcctcaggaatatttttctgtctgtgtactcagcctatctgtctg 1767
Oy 841 aatggctctcagaaatctatgtctctctcatctctagtgccccctatgacaactac 900
1768 aatggctctcagaaatctatgtctctctcatctctagtgccccctatgacaactac 1827
Oy 901 actgaaacaagattatatacaatcatgtcgtaccctaaagccccacaacaaagataccat 960
1828 actgaaacaagattatatacaatcatgtcgtaccctaaagccccacaacaaagataccat 1887
Oy 961 cttctctttgttatcagagcaggaatgtctagcagactagatcagatcgttgccagatc 1020
1888 cttctctttgttatcagagcaggaatgtctagcagactagatcagatcgttgccagatc 1947
Oy 1021 acaacccctactcagttctactacaacactatctcaagaataatcgtgacatgtgacaag 1080
1948 acaacccctactcagttctactacaacactatctcaagaataatcgtgacatgtgacaag 2007
Oy 1081 gtactgacatccctggtcaactctgcaagatcaacttaactccctcagcagtagtccct 1140
2008 gtactgacatccctggtcaactctgcaagatcaacttaactccctcagcagtagtccct 2067
Oy 1141 caaaatcgaaagagctttagactgtctaaacgcaagaagggggaagcctgtttatttca 1200
2068 caaaatcgaaagagctttagactgtctaaacgcaagaagggggaagcctgtttatttca 2127
Oy 1201 ggaagaagaacgctgttattatgttaataatccagaatgttcaactgagagaagttaaaga 1260
2128 ggaagaagaacgctgttattatgttaataatccagaatgttcaactgagagaagttaaaga 2187
Oy 1261 attcgagatcgaaatcaatgttagagcagaagagcttcaaaaacccgaacgctgggctc 1320
2188 attcgagatcgaaatcaatgttagagcagaagagcttcaaaaacccgaacgctgggctc 2247
Oy 1321 ctcagacaaatggaatgctctgggttctccctctcttagagacccctcagcagcttaattg 1380
2248 ctcagacaaatggaatgctctgggttctccctctcttagagacccctcagcagcttaattg 2307
Oy 1381 ttaactcctcttggaacccctgttaactccttgaactcctgttgaagtgtctctccagaatc 1440
2308 ttaactcctcttggaacccctgttaactccttgaactcctgttgaagtgtctctccagaatc 2367
Oy 1441 gaaactgttaaaagctacagatggtcttacaatggaacccca 1481
2368 gaaactgttaaaagctacagatggtcttacaatggaacccca 2408

RESULT 7
AAH20070
ID AAH20070 standard; DNA: 1617 BP.
XX AAH20070;
AC
XX
DT 08-AUG-2001 (first entry)

XX HERV-W envelope protein G coding sequence.
DE
XX
XX Human endogenous retrovirus: HERV-W; HERV: chromosome 7; env protein;
KW surface antigen; multiple sclerosis-related superantigen; vaccine;
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW antisense-therapy; autoimmune disorder; ds.
XX
OS Human endogenous retrovirus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..1617
FT /tag= a
FT /product= "HERV-W envelope protein G"
PM
XX
XX MO2001J1021-A1.
PD 03-MAY-2001.
XX
XX 30-OCT-2000; 2000WO-EP10659.
PF
XX 28-OCT-1999; 99EP-0402690.
PR
XX (UYGE-) UNIV GENEVE.
PA
XX
XX Conrad B, Mach B;
PI
XX WPI: 2001-316336/73.
DR P-PSDB: AAB75138.
XX
XX
PT New human retrovirus HERV-W ENV proteins/peptides having superantigen
PS activity useful for diagnosing and treating multiple sclerosis -
PS Claim 13; Fig 10; 94pp; English.
XX
XX On the basis of the PBS t-RNA motif used for the classification of human
CC endogenous retrovirus (HERVs) the full length endogenous provirus which
CC was located on the long arm of human chromosome 7 (7q21-22) has been
CC designated HERV-W. The present invention describes proteins or peptides
CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC have neuroprotective activity, and can be used in: vaccines; antisense-
CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC disorders. (I) are also useful for identifying substances (and optionally
CC recovering) capable of binding to a retroviral superantigen associated
CC with MS, substances capable of blocking SAG activity and substances
CC capable of blocking transcription or translation of HERV-W retroviral
CC superantigen. A protein or peptide derived from (I), modified to be
CC devoid of SAG activity and being capable of generating an immune response
CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid.
CC molecules encoding (I) are useful as vaccines against MS. Substances
CC capable of blocking SAG activity, capable of binding to a retroviral
CC superantigen associated with MS, or capable of blocking transcription or
CC translation of HERV-W retroviral superantigen for use in treating or
CC preventing MS, obtained using (I) are useful for the treatment and
CC prevention of MS. (I) and nucleic acids encoding them are useful for
CC diagnosing autoimmune disease. The present sequence encodes the
CC specifically claimed envelope protein of HERV-W designated G.
XX
XX
SO Sequence 1617 BP; 442 A; 452 C; 296 G; 427 T; 0 other;

Query Match 90.1%; Score 133.8; DB 22; Length 1617;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Oy 1 atggcctccctatcatacttctcttactctgtctcttacccttcgctcact 60
|||||
Db 1 atggcctccctatcatatttctcttactctgtctcttacccttcactcact 60
|||||
Oy 61 gacccctcatatgtctgttacaacacagtagctcccttacaagaattctatgaaga 120
|||||

Db 61 gaacccctccatgctgtatgacagtagtccctaccagaagtctctatggaga 120
 QY 121 acgagcttccctgaaatatgtagcccatcatatagaggttatctaagggaatcc 180
 Db 121 atgacagctccggaatatatgtagcccatcatatagaggttatctaagggaatcc 180
 QY 181 acctactgcccacacccatgccccgaactgctataactctgccaactcttgatg 240
 Db 181 acctactgcccacacccatgccccgaactgctataactctgccaactcttgatg 240
 QY 241 catgcaaatatcatatatttgacagaggaataatctcctagtctcttgagagct 300
 Db 241 catgcaaatatcatatatttgacagaggaataatctcctagtctcttgagagct 300
 QY 301 gggagccactgctgttgacttaactcaccatcacagtagtctctgaggggagat 360
 Db 301 gggagccactgctgttgacttaactcaccatcacagtagtctctgaggggagat 360
 QY 361 caaggtcaggcaagagaaaaaagaagaaagaaagaaagaaagaaagaaagaaag 420
 Db 361 caaggtcaggcaagagaaaaaagaagaaagaaagaaagaaagaaagaaagaaag 420
 QY 421 agacccctgagccctcacaagaagactagttctcacaactacatgaacccctgac 480
 Db 421 agacccctgagccctcacaagaagactagttctcacaactacatgaacccctgac 480
 QY 481 catactgctgtgtagcatttaatacaccctcactgctgctcactgaggtctagcc 540
 Db 481 catactgctgtgtagcatttaatacaccctcactgctgctcactgaggtctagcc 540
 QY 541 caaaacccctactactgtagttagtgcctccctgacactcagggccatcatatc 600
 Db 541 caaaacccctactactgtagttagtgcctccctgacactcagggccatcatatc 600
 QY 601 cctgttctcgaacaatggaacactcagcagcaagaataaacacactccgtttgta 660
 Db 601 cctgttctcgaacaatggaacactcagcagcaagaataaacacactccgtttgta 660
 QY 661 ggaacctctgttcccaatctggaataacccatcctcaaacctcactctgttaaat 720
 Db 661 ggaacctctgttcccaatctggaataacccatcctcaaacctcactctgttaaat 720
 QY 721 agcaatactatagacaacaacagctcccaatgcatcagtggtgtaaacctccacaga 780
 Db 721 agcaatactatagacaacaacagctcccaatgcatcagtggtgtaaacctccacaga 780
 QY 781 atagctgctacccctcagaataattttgtctgtgtacccacgtcatcatgtttg 840
 Db 781 atagctgctacccctcagaataattttgtctgtgtacccacgtcatcatgtttg 840
 QY 841 aatgagctcctcagaataattgtctcctcactcactcactcactcactcactcact 900
 Db 841 aatgagctcctcagaataattgtctcctcactcactcactcactcactcactcact 900
 QY 901 actgaaacaagattatatacaatcgtctactcactcactcactcactcactcact 960
 Db 901 actgaaacaagattatatacaatcgtctactcactcactcactcactcactcact 960
 QY 961 ctctccttcttatacagagcagagtgctgagcagactagctgagcattgagcatt 1020
 Db 961 ctctccttcttatacagagcagagtgctgagcagactagctgagcattgagcatt 1020
 QY 1021 acaacccctactcaggtctcactcactcactcactcactcactcactcactcact 1080
 Db 1021 acaacccctactcaggtctcactcactcactcactcactcactcactcactcact 1080
 QY 1081 gtcactgactcctcgtcactcgtcagagatcaactaactcctcactgagcagtagtct 1140
 Db 1081 gtcgagcactcctcgtcactcgtcagagatcaactaactcctcactgagcagtagtct 1140
 QY 1141 caaatcgaagagctttagactgctacacgccaagaagagggagacgtttatttta 1200
 Db 1141 caaatcgaagagctttagactgctacacgccaagaagagggagacgtttatttta 1200

QY 1201 ggaagaagacgtgttatgttaataatcaatccagaattgtcaactggaagaattaaaga 1260
 Db 1201 ggggaagaagacgtgttatgttaataatcaatccagaattgtcaactggaagaattaaaga 1260
 QY 1261 atcagagatcgaatataatgtagagcagagagcttcaaacacccagacgtgggctc 1320
 Db 1261 atcagagatcgaatataatgtagagcagagagcttcaaacacccagacgtgggctc 1320
 QY 1321 ctacgccaatgagatgacctggtctccctctcttagagccctcagacgtcataatg 1380
 Db 1321 ctacgccaatgagatgacctggtctccctctcttagagccctcagacgtcataatg 1380
 QY 1381 ttactcctcttgagcccttatcttcaactcctgttgaattgtctctccagaat 1440
 Db 1381 ctactcctcttgagcccttatcttcaactcctgttgaattgtctctccagaat 1440
 QY 1441 gaagctgtaaagctacagatggtcttacaagaatggaaccca 1481
 Db 1441 gaagctgtaaagctacagatggtcttacaagaatggaaccca 1481
 RESULT 8
 AAF55630 standard; DNA; 2781 BP.
 ID AAF55630:
 AC AAF55630:
 DT 29-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
 XX Envelope protein; HERV; syncytia formation; placental development.
 KW syncytia; cancer; cell adhesion; ss.
 XX
 OS Human endogenous retrovirus.
 XX
 FH Key Location/Qualifiers
 FT CDS 762..2378
 FT /*tag= "a
 FT /*product= "envelope protein"
 XX
 PD WO200116171-A1.
 XX
 PN 08-MAR-2001.
 XX
 PF 01-SEP-2000; 2000MO-FR02429.
 XX
 PR 01-SEP-1999; 99FR-0011141.
 PR 15-SEP-1999; 99FR-0011793.
 XX
 PA (INMR) BIO MERIEUX.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 XX
 PI Maillet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
 DR WPI: 2001-226676/23.
 DR P-PSDB; AAB67652.
 XX
 PT Detecting expression of human endogenous retrovirus envelope protein in
 XX cells of a tissue or culture, from its ability to induce syncytia -
 PS Disclosure: Page 44-45: 57pp; French.
 XX
 CC The present sequence encodes a human endogenous retrovirus envelope
 CC protein. The specification describes a method for detecting expression
 CC of an envelope protein from a human endogenous retrovirus (HERV), in
 CC cells, of a tissue or culture. The method comprises detecting syncytia
 CC formation due to the fusogenic properties of the envelope protein.
 CC Envelope polypeptides and polynucleotides are used to produce
 CC therapeutic or prophylactic compositions, particularly for treatment of
 CC cancer, to correct defects in placental development (or other natural
 CC formation of other types of syncytia), and to promote adhesion of cells

RESULT 10
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 BP.
XX
AC AAK25665;
XX
DT 21-MAY-1999 (first entry)
DE Complete human endogenous retrovirus W genome.
XX
KM Clone: human endogenous retrovirus; genome; autoimmune disease;
KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
OS Human endogenous retrovirus.
XX
PN WO9902696-A1.
XX
XX 21-JAN-1999.
PF 06-JUL-1998; 98WO-FR01442.
XX
PR 07-JUL-1997; 97FR-0008815.
XX
PA (INMR) BIO MERIEUX.
PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
XX
DR WPI: 1999-120897/10.
XX
PT New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
PS Claim 1; Page 71-74; 106pp; French.

XX
XX This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin- dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.
CC
XX
SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 89.3%; Score 1322.6; DB 20; Length 7582;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1364; Conservative 24; Mismatches 93; Indels 0; Gaps 0;

QY 1 atggccctcccttatcatcttctcttcttcttcttcttcttcccttgcctcact 60
DB 5581 atggscctcccttatcatcttctcttcttcttcttcttcttcccttgcctcact 5640
QY 61 gaacccccccatgctgctgaacaacagtagctcccttaccagaagttctatgaa 120
DB 5641 gaacccccccatgctgctgctgaacaacagtagctcccttaccagaagttctatgaa 5700
QY 121 acgagctcctctgtaaatatgtatgcccacatcatatagaggtttactaaggaaatcc 180
DB 5701 atgcaagctcccgaaatatgtatgcccacatcatatagaggtttactaaggaaatcc 5760
QY 181 accttaccgcccacacacatatagtcccgcaactgtataacttgcaccttttgacg 240
DB 5761 accttaccgcccacacacatatagtcccgcaactgtataacttgcaccttttgacg 5820
QY 241 catgcaaatcaactatattgagcaggaagaaatgatataatccctagttgcctgaggaact 300
DB 5821 catgcaaatcaactatattgagcaggaagaaatgatataatccctagttgcctgaggaact 5880

QY 301 ggagccactgctcgttggacttactcaaccacacacagatgctgtaggggtggaatt 360
DB 5881 ggagtcactgctcgttggacttactcaaccacacacagatgctgtaggggtggaatt 5940
QY 361 caagtcagcaggaagaaacaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 420
DB 5941 caagtcagcaggaagaaacaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 6000
QY 421 agcaaccctagccctacacaaagagtagttctctcaaaactacatgaaacccctgtaac 480
DB 6001 ggcaaccctagccctacacaaagagtagttctctcaaaactacatgaaacccctgtaac 6060
QY 481 catactgcctggtgagcctatttaatacaccctcaactcgcgtccatgaggtcagcc 540
DB 6061 catactgcctggtgagcctatttaatacaccctcaactcgcgtccatgaggtcagcc 6120
QY 541 caaaaccctactaactggtgagctggtgcccctccctgcaactttaggcacatacttaac 600
DB 6121 caaaaccctactaactggtgagctggtgcccctccctgcaactttaggcacatacttaac 6180
QY 601 cctgttcctgaacaatggaacaacttcaagacagaataaacaacacactcgttttagta 660
DB 6181 cctgttaccctgaacaatggaacaacttcaagacagaataaacaacacactcgttttagta 6240
QY 661 ggacctctgttctccaatctggaatbaaccacacacacacacacacacacacacacac 720
DB 6241 ggacctctgttctccaatctggaatbaaccacacacacacacacacacacacacacac 6300
QY 721 agcaatactatgac 780
DB 6301 agcaatactatgac 6360
QY 781 atagctgctcctccctcagaatatttttctgctgctgagctcagcctcatattgttg 840
DB 6361 atagctgctcctccctcagaatatttttctgctgctgagctcagcctcatattgttg 6420
QY 841 aatgctcttcaagaatctatgcttctcctcactcacttcttagtgccccctatgacacatc 900
DB 6421 aatgctcttcaagaatctatgcttctcctcactcacttcttagtgccccctatgacacatc 6480
QY 901 actgacaagaatttatcatatgcttctgctgctgagctcagcctcagcctcagcctcagc 960
DB 6481 actgacaagaatttatcatatgcttctgctgctgagctcagcctcagcctcagcctcagc 6540
QY 961 ctctcttctgtatcagaagcagagtgctcagcagcagcagcagcagcagcagcagcagc 1020
DB 6541 ctctcttctgtatcagaagcagagtgctcagcagcagcagcagcagcagcagcagcagc 6600
QY 1021 acaacctactcactcacttctactacaaactatctcaagaataatagtgatgaaacag 1080
DB 6601 acaacctactcactcacttctactacaaactatctcaagaataatagtgatgaaacag 6660
QY 1081 gtcactgactcctcctgctcacttgcagagtcacataactcctcctcagcagcagtaagcct 1140
DB 6661 gtcactgactcctcctgctcacttgcagagtcacataactcctcctcagcagcagtaagcct 6720
QY 1141 caaatcgaagagctttagacttgcacacgccaagaaggggggaacccgtttatttta 1200
DB 6721 craaatcgaagagctttagacttgcacacgccaagaaggggggaacccgtttatttta 6780
QY 1201 ggaagaagaagctgtatattatgttaactcaatccagaattgtcaactgagaagctaaaga 1260
DB 6781 ggggaagaagaagctgtatattatgttaactcaatccagaattgtcaactgagaagctaaaga 6840
QY 1261 attcgagatcgaaatcaaatgtagagcagaggggtttcaaaacacgaaagcggggccct 1320
DB 6841 attcgagatcgaaatcaaatgtagagcagaggggtttcaaaacacgaaagcggggccct 6900
QY 1321 ctcacgaatagatgccctgggttctccctcctttagagccctcagcagccttaaatg 1380
DB 6901 ctcacgaatagatgccctgggttctccctcctttagagccctcagcagccttaaatg 6960
QY 1381 ttactcctcttggaacctgtactttaaactcctctgttaagttgtctctccagaatt 1440

Db	6961	ctaccctcccttggaccctgtacattcttacccttctgttaactctgtctctccagaatc	7020
Qy	1441	gaagctgtlaagctacagatgctctacaatggaaccca	1481
		:	
Db	7021	gaagctgttaactactacaatgtgagcccaagatgcaatccaa	7061
RESULT	11		
	AAH20069		
ID	AAH20069	standard; DNA; 2782 bp.	
XX			
AC	AAH20069;		
XX			
DT	08-AUG-2001	(first entry)	
XX			
DE	HERV-W envelope protein G encoding nucleic acid.		
XX			
KW	Human endogenous retrovirus: HERV-W; HERV; chromosome 7; env protein;		
KX	envelope protein; multiple sclerosis-related superantigen; vaccine;		
XX	surface antigen; transmembrane; multiple sclerosis; neuroprotective;		
KW	antisense therapy; autoimmune disorder; ds.		


```

QY 601 ccttctcgaacaatggaacaacttcagcacaagaataaaccacttcgttttagta 660
DB 1363 ccgttacttgaaacaatggaacaacttcagcacaagaataaaccacttcgttttagta 1422
QY 661 ggaactctgtttccaactctggaataaaccatacctcaaacctcactctgttataatt 720
DB 1423 ggaactctgtttccaactctggaataaaccatacctcaaacctcactctgttataatt 1482
QY 721 agcaatactatagaacaacaacagctcccaatgcatcagggggtgaataacacctccacacga 780
DB 1483 agcaatactatagaacaacaacagctcccaatgcatcagggggtgaataacacctccacacga 1542
QY 781 ataagtcgtcctacccaagaataatttttctgtctgttaccctagacctatcatgttttg 840
DB 1543 ataagtcgtcctacccaagaataatttttctgtctgttaccctagacctatcatgttttg 1602
QY 841 aatggtctctcagaatactatgtgtctccctcattcttaagtgtccctatgacacatcac 900
DB 1603 aatggtctctcagaatactatgtgtctccctcattcttaagtgtccctatgacacatcac 1662
QY 901 actgaaacaagattatatacaatcagtcgttaccttaagccccaacaagaaggtaccatt 960
DB 1663 actgaaacaagattatatacaatcagtcgttaccttaagccccaacaagaaggtaccatt 1722
QY 961 ctctcctttttatcagagcagaggtgtcagcagactaagttactgtgcatgtgcagcatc 1020
DB 1723 ctctcctttttatcagagcagaggtgtcagcagactaagttactgtgcatgtgcagcatc 1782
QY 1021 acaacctctactcagttcttactatacaaatcctcaagaataaattgtgtacatgtgaacag 1080
DB 1783 acaacctctactcagttcttactatacaaatcctcaagaataaattgtgtacatgtgaacag 1842
QY 1081 gtcaactgacctccctggtgcaccttgcagaatcaacttaacttccctagcagcagtgctct 1140
DB 1843 gtccgcagactccctggtgcaccttgcagaatcaacttaacttccctagcagcagtgctct 1902
QY 1141 caaatatggaagagctttagacttcttaaccgcacaagaagggggaacctgtttatttta 1200
DB 1903 cgaatctggaagagctttagacttcttaaccgcagagagggggaacctgtttatttta 1962
QY 1201 ggagaagaacgctgttattatgttaatcaatccagaattgtcactgtgagaagaattaaaga 1260
DB 1963 ggagaagaacgctgttattatgttaatcaatccagaattgtcactgtgagaagaattaaaga 2022
QY 1261 attggaatcgaaatgaatgagagcagagagcttcaaaacaccagcagctgtggcctc 1320
DB 2023 attcagatcgaaatgaatgagagcagagagcttcaaaacaccagcagctgtggcctc 2082
QY 1321 ctcaagcaatgtagctgtggtctccctctcttaagagacctctagcagcgtctaatttg 1380
DB 2083 ctcaagcaatgtagctgtggtctccctctcttaagagacctctagcagcgtctaatttg 2142
QY 1381 ttactctcttttgaaacctgtatcttaacctcctgtttaaagttgtctctccagaatt 1440
DB 2143 ttactctcttttgaaacctgtatcttaacctcctgtttaaagttgtctctccagaatt 2202
QY 1441 gaaactgtgaaagctacagatgtcttacaagtgaacccca 1481
DB 2203 gaaactgtgaaagctacagatgtcttacaagtgaacccca 2243

```

RESULT 13

AAA59211
ID AAA59211 standard; DNA; 2782 BP.

XX AAA59211:

XX 07-NOV-2000 (first entry)

XX 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.

XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;

```

KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX Homo sapiens.
OS
XX MO200043521-A2.
PN
XX 27-JUL-2000.
PD
XX 21-JAN-2000; 2000MO-FR00144.
PF
XX 21-JAN-1999; 99PR-0000888.
PR
XX (INMR ) BIO MERIEUX.
PA
PI Paranhos-Baccala G, Mallet F, Voliset C;
DR WPI: 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy,
PT contains at least part of the gag gene
XX
XX Disclosure; Page 46-47; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is
CC integrated into the human genome. The fragment is originally derived
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
XX Sequence 2782 BP: 741 A; 767 C; 565 G; 709 T; 0 other:
SQ

```

Query Match 88.2%; Score 1306.6; DB 21; Length 2782;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1372; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

QY 1 atggccctccctatcatcttctcttactgttctcttccctccctcctcact 60
DB 763 atgggctcccttatacatatttctctgtatgttcttcaacctgtttcaacctcact 822
QY 61 gcaacccctcatgtctgtacacaacagtagctcccttaccgaagttctatgaaga 120
DB 823 gcaacccctcatgtctgtacacaacagtagctcccttaccgaagttctctatgaga 882
QY 121 acgggctctctggaataatgatgcccacatagatgaattatctaaaggaaacctc 180
DB 883 atgcagcgttccgggaatatgatgcccacatagatgaattatctctaaaggaaacctc 942
QY 181 accttcatgcccacacacatagcccgcgaactgctataactctgccaactcttgatg 240
DB 943 accttcatgcccacacacatagcccgcgaactgctataactctgccaactcttgatg 1002
QY 241 catgcaatcatattatgacaggggaataatgaaatctctagtgtctcgtgaagact 300
DB 1003 catgcaatcatattatgacaggggaataatgaaatctctagtgtctcgtgaagact 1062
QY 301 ggaagcactgtctgttgaacttacttccacataccatagctcgtgaagggtgaatt 360
DB 1063 ggaagcactgtctgttgaacttacttccacataccatagctcgtgaagggtgaatt 1122
QY 361 caaggtcagcaaggaataaacaagttaaaggagcaatccccaactgaccggggaat 420
DB 1123 caagatcagcaaggaataaacaagttaaaggagcaatccccaactgaccggggaat 1182
QY 421 agcaccctagccctacaaaggactgtctctcaaaactacatgaaacctccgtacc 480

```

Db 1183 ggcacccctgaagcccccacaaaggactagatctctcaaaactatcgtaaaacctcgttac 1242
Qy 481 cactactgcgcctggtgagcctatttaataccacccctcactcggctccatgaggtctcagcc 540
Db 1243 cactacgcgcctgttaagcctatttaataccacccctcactcggctccatgaggtctcagcc 1302
Qy 541 caaaacccctactaacgtgttgatgtgctcccccctgcaacttcagagccatataatccaac 600
Db 1303 caaaacccctactaacgtgttgatgtgctcccccctgcaacttcagagccatataatccaac 1362
Qy 601 cctgttcctgaacaatgtagaacacttcagacagaaataaacaccactccgttttagta 660
Db 1363 cctgttcctgaacaatgtagaacacttcagacagaaataaacaccactccgttttagta 1422
Qy 661 ggaacccctgttctccaacttggaataaacccatccctcaaacctcactcgttgaaatct 720
Db 1423 ggaacccctgttctccaacttggaataaacccatccctcaaacctcactcgttgaaatct 1482
Qy 721 agcaataactatagacaacaacagctcccaatgcataggttggttaaacctccacacga 780
Db 1483 agcaataactatagacaacaacagctcccaatgcataggttggttaaacctccacacaa 1542
Qy 781 atagctgtcctacccccacagaatatctttgtctgtgtagctcagccatcatgtgttg 840
Db 1543 atagctgtcctacccccacagaatatctttgtctgtgtagctcagccatcatgtgttg 1602
Qy 841 aatgtgctctcagaaatctatgtgtctcctcctcatctctatgtgtgccccctatgacatctac 900
Db 1603 aatgtgctctcagaaatctatgtgtctcctcctcatctctatgtgtgccccctatgacatctac 1662
Qy 901 actgtaacaaagattataacaatactgtcgtacttaagccccacaaacaaagaatacccat 960
Db 1663 actgtaacaaagattataacaatactgtcgtacttaagccccacaaacaaagaatacccat 1722
Qy 961 cctcctcttgatctacgaacagagagtgctcagcagactagctgacatgctgacatgac 1020
Db 1723 cctcctcttgatctacgaacagagagtgctcagcagactagctgacatgctgacatgac 1782
Qy 1021 acaacccctactcagttctactacaacaactatctcaagaaataatgtgtacatggaacag 1080
Db 1783 acaacccctactcagttctactacaacaactatctcaagaaataatgtgtacatggaacag 1842
Qy 1081 gtcacactgactcctgtgtacacttgcaagttcaacttaactcctctagcagatagtcctt 1140
Db 1843 gtcacactgactcctgtgtgtacacttgcaagttcaacttaactcctctagcagatagtcctt 1902
Qy 1141 caaaatcgaagaagctttaactgtcctaacgcgaacaaagaagggggaacccgtttatttta 1200
Db 1903 caaaatcgaagaagctttaactgtcctaacgcgtgagaggggggaacccgtttatttta 1962
Qy 1201 ggaagaagaagcgtgtatatagtttaataccagaatgtgtcactgtgagaagttaagaa 1260
Db 1963 gggggaagaagcgtgtatatagtttaataccagaatgtgtcactgtgagaagttaagaa 2022
Qy 1261 atctgagatcgaataacatagtagagcagaggttcaaaaacacgaacgcctggggccct 1320
Db 2023 atctcgaatcgaataacatagtagagcagaggttcaaaaacacgaacgcctggggccct 2082
Qy 1321 ctcagagcaatgtagtcgccctgggtctcccccctcttagagccctcagcacttaatatgt 1380
Db 2083 ctcagagcaatgtagtcgccctgggtctcccccctcttagagccctcagcacttaatatgt 2142
Qy 1381 ttaactcctcttggaacccgtatctttaacctcctctgttaagttgtctctccagaatt 1440
Db 2143 ctactcctcttggaacccgtatcttggaacccctctgttaactgtctctctccagaatt 2202
Qy 1441 ggaagctgtaagaactcagaatggtcttatacaaatgtaacccca 1481
Db 2203 ggaagctgtaagaactcagaatggtcagatgcagttccaa 2243

RESULT 14
ABAA5822

ID ABAA5822 standard: DNA; 1894 BP.
XX
AC ABAA5822;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #4517.
XX
KW Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer; ss.
OS Homo sapiens.
PN WO200157271-A2.
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
DR
XX
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PS breast, comprises number of single exon nucleic acid probes -
PS
PS
PS Claim 1: SEQ ID NO 4517; 327pp + sequence listing; English.
XX
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Bt 474 cells. The method involves contacting
CC the probes with a collection of detectably labeled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 other;

Query Match 76.4%; Score 1131.6; DB 22; Length 1894;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1218; Conservative 0; Mismatches 84; Indels 16; Gaps 2;

Qy 164 tatctaaaggaaagccactctactgtgcccacacatattgcccgcagactgtataact 223
Db 1 tatctaaaggaaagccactctactgtgcccacacatattgcccgcagactgtataact 60
Qy 224 ctgcacactcttgatgcatgtaagaatactatattgtgacagaggaatagtataatccta 283
Db 61 ctgcacactcttgatgcatgtaagaatactatattgtgacagaggaatagtataatccta 120

QY 284 gtgtctcgtgaggaactgtaggcacgtgtctgttggaacttaactcaaccatacagatagt 343
DB 121 gtgtctcgtgaggaactgtaggcacgtgtctgttggaacttaactcaaccatacagatagt 167
QY 344 ctgctggtggtgtgaatccaagtcacgaagagaagaaacagtaaggaagcaatccccc 403
DB 168 ctgctggtggtgtgaatccaagtcacgaagagaagaaacagtaaggaagcaatccccc 227
QY 404 aactgagccgggagacatagacccctagccctcaacaaagacatagctctctcaaaaactac 463
DB 228 aactgagccgggagacatagacccctagccctcaacaaagacatagctctctcaaaaactac 287
QY 464 atgaaacccctcgttaaccatactcgtgtgtgagcctatttaatacaccatacctcgtgc 523
DB 288 atgaaacccctcgttaaccatactcgtgtgtgagcctatttaatacaccatacctcgtgc 347
QY 524 tccatgaggtctcagcccaaacccatacctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 583
DB 348 tccatgaggtctcagcccaaacccatacctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 407
QY 584 ggcatacatctcaatccctgttccctgaacaatggaacaacttcagacagagaagaaataaca 643
DB 408 ggcatacatctcaatccctgttccctgaacaatggaacaacttcagacagagaagaaataaca 467
QY 644 ccacttcgtttagtaggaacctgtgttcccaatctggaataaccatacctcaaac 703
DB 468 ccacttcgtttagtaggaacctgtgttcccaatctggaataaccatacctcaaac 524
QY 704 tccactgtgtgaaatattagaataactataacacacacacagctcccaatgcatcagtggtg 763
DB 525 tccactgtgtgaaatattagaataactataacacacacacagctcccaatgcatcagtggtg 584
QY 764 taacacccctccacacgaatagctgtcctaccctcagaagaatattttgtgtgtgtgtgtgt 823
DB 585 taactcctccacacgaatagctgtcctaccctcagaagaatattttgtgtgtgtgtgtgtgt 644
QY 824 cagactatcatgtttagaagtcctcagaatcatatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 883
DB 645 cagactatcatgtttagaagtcctcagaatcatatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 704
QY 884 cccctatgaccatctacacacgaagaattatatacaatctgtgtgtgtgtgtgtgtgtgtgtgt 943
DB 705 cccctatgaccatctacacacgaagaattatatacaatctgtgtgtgtgtgtgtgtgtgtgtgt 764
QY 944 acaaaagagacccatctctctgtttagtcagagagagtcgtgagagagagagagagagagag 1003
DB 765 acaaaagagacccatctctctgtttagtcagagagagtcgtgagagagagagagagagagag 824
QY 1004 ctgagcatgtgagatcac 1063
DB 825 ctgagcatgtgagatcac 884
QY 1064 atgtgtgacatggaacagtgctgt 1123
DB 885 atgtgtgacatggaacagtgctgt 944
QY 1124 tagcagcagtagtcctcacaac 1183
DB 945 tagcagcagtagtcctcacaac 1004
QY 1184 gaacctgttatatttttagagaagaacgctgttatatttttagaatacaatcagaatgttca 1243
DB 1005 gaacctgttatatttttagagaagaacgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1064
QY 1244 ctgagaagaatgtaagaatcgcagatcgaaatgtaagaatgtaagaatgtaagaatgtaagaat 1303
DB 1065 ctgagaagaatgtaagaatcgcagatcgaaatgtaagaatgtaagaatgtaagaatgtaagaat 1124
QY 1304 ccgaacgctgtggt 1363
DB 1125 ctgagacccctgt 1184
QY 1364 tagcagcctataatgttactcctctgtgagccgttatctttaacctctgtttagt 1423

DB 1185 tagcagcctataatgttactcctctgtgagccctgtatctttaacctctgtttagt 1244
QY 1424 ttgtctctccagaatggaagctgtgaagctacagatgtgtcttaacatggaaccaca 1481
DB 1245 ttgtctctccagaatggaagctgtgaagctacagatgtgtcttaacatggaaccaca 1302

RESULT 15
ID ABA56337 standard; DNA; 1894 BP.
XX ABA56337;
AC ABA56337;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #4642.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN W0200157277-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001MO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human foetal liver -
XX
PS Claim 1; SEQ ID NO 4642; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 other;

Query Match 76.4%; Score 1131.6; DB 22; Length 1894;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1218; Conservative 0; Mismatches 84; Indels 16; Gaps 2;

QY 164 tatctaaaggaaactcacctcactgtgcccacacacacacacacacacacacacacacacacacac 223
DB 1 tatctaaaggaaacccacactcactgtgcccacacacacacacacacacacacacacacacacacac 60
QY 224 ctgacacactctgtgacatgac 283
DB 61 ctgacacactctgtgacatgac 120
QY 284 gtgtctcgtgaggaactgtaggcacgtgtctgttggaacttaactcaaccatacagatagt 343

Db 121 gttgttcctggaggaacttgg-----ggactcaactctctacacagtaagt 167
Qy 344 cggatcggggttggaattcgaaggttcagggcaagaagaacaaagtaaaggagcaatctccc 403
Db 168 cggatcggggttggaattcgaaggttcagggcaagaagaagaacaaagtaaaggagtaactccc 227
Qy 404 aactgaaccggggaacatagcaacccttagccctacaaagaagtaagtctctcaaaactac 463
Db 228 aactgaaccggggaacatagcaacccttagccctacaaagaagtaagtctctcaaaactac 287
Qy 464 atgaaaccctcctgtaacccttagctgctgtaagcctatttaaccactcactcggc 523
Db 288 atgaaaccctcctgtaacccttagctgctgtaagcctatttaaccactcactcggc 347
Qy 524 tcacatgaagctcgaagcccaaaccttaactagtgtggaatgctgcctccctcgcaactca 583
Db 348 tcacatgaagctcgaagcccaaaccttaactagtgtggaatgctgcctccctcgcaactca 407
Qy 584 ggcacataactcaatccctgttccctgaaacaaatgaaacaaacttcagcacagaataaaca 643
Db 408 ggcacataactcaatccctgttccctgaaacaaatgaaacaaacttcagcacagaataaaca 467
Qy 644 ccaacttcggttttagtaagacctctgtttccaatctggaataaaccatacctcaaac 703
Db 468 ccaacttcggttttagtaagacctctgtttccaatctggaataaaccatacctcaaac 524
Qy 704 tcaactgtgttaaatcttaacataactatagacacaaacagccccaatgcacagtgagg 763
Db 525 tcaactgtgttaaatcttaacataactatagacacaaacagccccaatgcacagtgagg 584
Qy 764 taacacctcccaacgaatagctgcctacccctcaggaatattttgtcgtgtaacct 823
Db 585 taacacctcccaacgaatagctgcctacccctcaggaatattttgtcgtgtaacct 644
Qy 824 cagcctatcatgtttgtaatggctcttcagaatctatgtgctcctcctcatctagtg 883
Db 645 cagcctatcatgtttgtaatggctcttcagaatctatgtgctcctcctcatctagtg 704
Qy 884 cccctaagcaactcaacatgaagaatcttaacatcatgtcgtacctaagcccaaca 943
Db 705 cccctaagcaactcaacatgaagaatcttaacatcatgtcgtacctaagcccaaca 764
Qy 944 acaaaaagatcaccaatctctctgtgtatcagaagcaggaagtgcagcagactagta 1003
Db 765 acaaaaagatcaccaatctctctgtgtatcagaagcaggaagtgcagcagactagta 824
Qy 1004 ctggcatggcggtatcaaacctctactcaagttctactacaactatctcaagaataa 1063
Db 825 ctggcatggcggtatcaaacctctactcaagttctactacaactatctcaagaataa 884
Qy 1064 atggctgaactgaagcaagctcaactgcaactccctgcaacttcgaagatacaactccc 1123
Db 885 atggctgaactgaagcaagctcaactgcaactccctgcaacttcgaagatacaactccc 944
Qy 1124 tagcagcagtagtctctcaaaatcgaaagccttagacttgctaaacggcaaaagagg 1183
Db 945 tagcagcagtagtctctcaaaatcgaaagccttagacttgctaaacggcaaaagagg 1004
Qy 1184 gaacctgttatttttagagaagaagcgtgtatattgttaatcaatccagaattgtca 1243
Db 1005 gaacctgttatttttagagaagaagcgtgtatattgttaatcaatccagaattgtca 1064
Qy 1244 ctgagaagaatgaagaatctgaatcgaatacaatgttagagcagagagcctcaaaa 1303
Db 1065 ctgagaagaatgaagaatctgaatcgaatacaatgttagagcagagagcctcaaaa 1124
Qy 1304 ccgaacgcttgaggcctcctcagccaatggaatgcccctgggtctccctcttagagcctc 1363
Db 1125 ctggacccctgggctcctcagccaatggaatgcccctgggtctccctcttagagcctc 1184
Qy 1364 tagcagcctaaatattgttaactccctcttgagcctgtatcttaacctctgtttaagt 1423

Db 1185 tagcagcctaataatgttaactccctcttgagaccctgtatcttcaacctctgtttaagt 1244
Qy 1424 tggctcttcagaatltgaagctgtlaaagctacagatggcttaacaatatggaaccca 1481
Db 1245 tggctcttcagaatltgaagcagtaaacataaatcgaatcgttcttcaaatggaaccca 1302

Search completed: June 20, 2002, 09:43:22
Job time: 8532 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:43:49 ; Search time 4429.18 Seconds
(without alignments)
6997.277 Million cell updates/sec

Title: US-09-319-156A-9
Perfect score: 1481
Sequence: 1 atggccctccctatac.....gtctacaatggaaccca 1481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_pro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	1481	100.0	1481	6	AX001027	AX001027 Sequence
2	1481	100.0	1932	14	AF127228	AF127228 Multiple
3	1468.2	99.1	162579	9	AL390039	AL390039 Human DNA
4	1461.8	98.7	1629	6	AX036992	AX036992 Sequence
5	1461.8	98.7	1629	14	AF331500	AF331500 Multiple
6	1461.8	98.7	172918	2	AC023366	AC023366 Homo sapi
7	1335.4	90.2	2930	6	AX355872	AX355872 Sequence
8	1335.4	90.2	2930	6	AF208161	AF208161 Homo sapi
9	1333.8	90.2	2946	6	AR177269	AR177269 Sequence
10	1333.8	90.1	2055	6	AX007999	AX007999 Sequence
11	1333.8	90.1	2539	6	AX007978	AX007978 Sequence
12	1333.8	90.1	10499	6	AX007980	AX007980 Sequence
13	1333.8	90.1	56093	6	AX329572	AX329572 Sequence
14	1333.8	90.1	56093	6	HSAC000064	AC000064 Human BAC
15	1333.8	90.1	149194	9	AC007566	AC007566 Homo sapi
16	1332.2	90.0	1617	9	AF156963	AF156963 Homo sapi
17	1330.6	89.8	2781	6	AX092223	AX092223 Sequence
18	1330.6	89.8	2781	6	AF072506	AF072506 Homo sapi
19	1324.2	89.4	7582	6	AX000966	AX000966 Sequence
20	1324.2	89.4	7582	6	AX027480	AX027480 Sequence
21	1318	89.0	153444	2	AC040948	AC040948 Homo sapi
22	1318	89.0	205035	9	CNS00005	AL049870 Human chr
23	1306.6	88.2	2782	6	AX000962	AX000962 Sequence
24	1306.6	88.2	2782	6	AX027476	AX027476 Sequence
25	1306.6	88.2	2782	6	AX133396	AX133396 Sequence
26	1297.4	87.6	161571	2	AC092708	AC092708 Homo sapi
27	1297.4	87.6	166700	9	AC013759	AC013759 Homo sapi
28	1297.4	87.6	186911	2	AP001488	AP001488 Homo sapi
29	1294	87.4	172721	2	AC008121	AC008121 Homo sapi
30	1285	87.4	188919	2	AC009727	AC009727 Homo sapi
31	1285	86.8	162522	9	AC093839	AC093839 Homo sapi
32	1282	86.6	46340	6	AX008043	AX008043 Sequence
33	1282	86.6	177163	9	HS134E15	AL022067 Human DNA
34	1278.8	86.3	1799	6	AX007982	AX007982 Sequence
35	1276.2	86.2	188038	2	AC092510	AC092510 Papio cyn
36	1271.2	85.8	178203	9	AC008780	AC008780 Homo sapi
37	1261	85.1	158033	9	AC018926	AC018926 Homo sapi
38	1258.8	85.0	169462	2	AC024033	AC024033 Homo sapi
39	1256.2	84.8	182178	9	AC074198	AC074198 Homo sapi
40	1255.8	84.8	183680	9	AC098859	AC098859 Homo sapi
41	1251.6	84.5	113253	2	AC092356	AC092356 Homo sapi
42	1247.8	84.3	186723	2	AC027752	AC027752 Homo sapi
43	1242.2	83.9	183267	2	AC068447	AC068447 Homo sapi
44	1239	83.7	2784	6	AX007981	AX007981 Sequence
45	1239	83.7	135734	2	AC002346	AC002346 Homo sapi

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN
1	AX001027	Sequence 9 from Patent WO9902666.	AX001027	AX001027.1	GI:7241264	unidentified.	unclassified.	1 (bases 1 to 1481)	Ott,C. and Bedin,F.	RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES	Patent: WO 9902666-A 9 21-JAN-1999;	location/Qualifiers	412 a	
												1.1481	410 c	261 g
												/organism="unidentified"	398 t	
												/db_xref="taxon:32644"		

Query Match 100.0%; Score 1481; DB 6; Length 1481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 atgagcccccatacaactttctcttactgttctcttaaccccccttgctcaact 60
Db 1 atggcccccatacaactttctcttactgttctcttaaccccccttgctcaact 60
Oy 61 gaacccccccatgctgtgtgtacaaacagtagctccccccttaccagaagtttctatgaaga 120
Db 61 gaacccccccatgctgtgtgtacaaacagtagctccccccttaccagaagtttctatgaaga 120
Oy 121 acgagcgtcttcctggaataatlgatgcccacatataagaagtttatactaaaggaactcc 180
Db 121 acgagcgtcttcctggaataatlgatgcccacatataagaagtttatactaaaggaactcc 180
Oy 181 accttcaactgcccacacacatatagcccgaactgtctataactctgcaactcttgcaatg 240
Db 181 accttcaactgcccacacacatatagcccgaactgtctataactctgcaactcttgcaatg 240
Oy 241 catgtcaaatctcatctatgtgacaggaagaatgatatacctgaagttgtccctggaagactt 300
Db 241 catgtcaaatctcatctatgtgacaggaagaatgatatacctgaagttgtccctggaagactt 300
Oy 301 ggaagcaactgtctgtgtgacttacttcaacacagtaigtctgatatgaggggtggaatt 360
Db 301 ggaagcaactgtctgtgtgacttacttcaacacagtaigtctgatatgaggggtggaatt 360
Oy 361 cabagtcagcagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 420
Db 361 cabagtcagcagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 420
Oy 421 agcaacccccatacctacaaaggaagtagttctctcaaaactcaatgaaacccctcgtaac 480
Db 421 agcaacccccatacctacaaaggaagtagttctctcaaaactcaatgaaacccctcgtaac 480
Oy 481 catactgcgcgtgtgagcctatttaatacaccactcaactcggtctcatgaggtctcaagcc 540
Db 481 catactgcgcgtgtgagcctatttaatacaccactcaactcggtctcatgaggtctcaagcc 540
Oy 541 caaaacccctacaaactgtgtgagtgtcctccctcgacacttcaagccataacttcaatc 600
Db 541 caaaacccctacaaactgtgtgagtgtcctccctcgacacttcaagccataacttcaatc 600
Oy 601 cctgtctctgacaacatggaacaaacttcagcagaagaacaacacacacttcggttttagta 660
Db 601 cctgtctctgacaacatggaacaaacttcagcagaagaacaacacacacttcggttttagta 660
Oy 661 ggaacctctgtttccaatctggaataaaccatacctcaaaccttcaactgtgttaaatct 720
Db 661 ggaacctctgtttccaatctggaataaaccatacctcaaaccttcaactgtgttaaatct 720
Oy 721 agcaatatactatagaacaacaacagctcccaatgcatcagtggttaacacttcccacaaga 780
Db 721 agcaatatactatagaacaacaacagctcccaatgcatcagtggttaacacttcccacaaga 780
Oy 781 atagctcgtcctacccctcagaataatcttctgtctgtgtacactcagcctatcatgtgttg 840
Db 781 atagctcgtcctacccctcagaataatcttctgtctgtgtacactcagcctatcatgtgttg 840
Oy 841 aatggtctctcaagaactatgcttctcctcaatctctatgtgagccctatgacactac 900
Db 841 aatggtctctcaagaactatgcttctcctcaatctctatgtgagccctatgacactac 900
Oy 901 actgaacaagaatttatatacatatgtctactaaagcccaacaagaagaagttaccctt 960
Db 901 actgaacaagaatttatatacatatgtctactaaagcccaacaagaagaagttaccctt 960
Oy 961 ctctctcttctatcagaagcaggaatgcttagcagaacaggtaactgcatctgagcagatc 1020
Db 961 ctctctcttctatcagaagcaggaatgcttagcagaacaggtaactgcatctgagcagatc 1020

```

```

Oy 1021 acaacctctactcaagttctactacaaactatctcaagaataaactggtgacatggaacag 1080
Db 1021 acaacctctactcaagttctactacaaactatctcaagaataaactggtgacatggaacag 1080
Oy 1081 gtcactgactccctgtgtcaactctgcaagatacaacttaactccctgacagcagtaactct 1140
Db 1081 gtcactgactccctgtgtcaactctgcaagatacaacttaactccctgacagcagtaactct 1140
Oy 1141 caaaatcgaaagcgttgaactgttaacccgcaaaagaaggggaacactggttatctta 1200
Db 1141 caaaatcgaaagcgttgaactgttaacccgcaaaagaaggggaacactggttatctta 1200
Oy 1201 ggaagaagaagcgtgttataatgttaataatccagaatgtgctacgtgagaagaatlaaaga 1260
Db 1201 ggaagaagaagcgtgttataatgttaataatccagaatgtgctacgtgagaagaatlaaaga 1260
Oy 1261 atcgagatcgataatacaatgtagagcagaagagccttcaaaacacagacgttgggcttc 1320
Db 1261 atcgagatcgataatacaatgtagagcagaagagccttcaaaacacagacgttgggcttc 1320
Oy 1321 ctcaagcaaatgagatccctggtgtctccctcttaaggaacccctagagccttaatatgt 1380
Db 1321 ctcaagcaaatgagatccctggtgtctccctcttaaggaacccctagagccttaatatgt 1380
Oy 1381 ttactcctcttggacccctgatacttcaactcctgttgaagttgtctctccagaact 1440
Db 1381 ttactcctcttggacccctgatacttcaactcctgttgaagttgtctctccagaact 1440
Oy 1441 gaagctgtgaagctcaagatggtcttacaagaatggaaccca 1481
Db 1441 gaagctgtgaagctcaagatggtcttacaagaatggaaccca 1481

```

RESULT 2
 AF127228
 LOCUS AF127228 1932 bp mRNA linear VRL 11-NUC-1999
 DEFINITION Multiple sclerosis associated retrovirus element clone CL15
 ACCESSION AF127228
 VERSION AF127228.1 GI:5726293
 KEYWORDS
 ORGANISM multiple sclerosis associated retrovirus element.
 Viruses; Retroid viruses; Retroviridae.
 REFERENCE 1 (bases 1 to 1932)
 Ounanian-Pradel, F., Paranhos-Baccala, G., Bedin, F., Mallet, F., Mandrand, B. and Perron, H. Molecular cloning and characterization of MSRV-related sequences associated with retrovirus-like particles
 JOURNAL Virology 260 (1), 1-9 (1999)
 MEDLINE 99335590
 PUBMED 10405350
 REFERENCE 2 (bases 1 to 1932)
 Ounanian-Pradel, F., Paranhos-Baccala, G., Bedin, F., Mallet, F., Mandrand, B. and Perron, H. Direct Submission
 JOURNAL Submitted (10-FEB-1999) UMR103 CNRS, biomerieux, 46, allée D'Italie, Lyon 69007, France
 FEATURES
 source
 1. 1932
 Location/Qualifiers
 /organism="multiple sclerosis associated retrovirus element"
 /db_xref="taxon:89382"
 /clone="CL15"
 452..1932
 /note="similar to env but interrupted by premature stop codon; includes signal peptide region, and partial SU and TM regions"
 misc_difference 568
 /note="in clones CL15-2 and CL37-14; results in a

IMPORTANT: This sequence is not the entire insert of clone RP13-383K5. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP13-383K5 is at 1 in this sequence. The true left end of clone RP5-1126E12 is at 162480 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RP13-383K5 is from the library RP1-13.2 constructed by the group of Pictet de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

FEATURES

```
Source
1..162579
/organism="Homo sapiens"
/dbxref="taxon:9606"
/chromosome="X"
/map="q22.1-24"
/clone="RP13-383K5"
/clone_1lb="RP1-13.2"
complement(1..232)
/note="match: GSS: Em:A0036363"
16..354
/note="match: GSS: Em:A0504848"
475..644
/note="Alusg repeat: matches 134..303 of consensus"
742..919
/note="MLT1B repeat: matches 207..388 of consensus"
1112..1142
/note="LMC repeat: matches 327..356 of consensus"
1143..1456
/note="Alusx repeat: matches 1..312 of consensus"
1457..2938
/note="LMC repeat: matches 356..1897 of consensus"
2922..3739
/note="L1PA2 repeat: matches 5329..6146 of consensus"
3844..4087
/note="L1M4 repeat: matches 1..251 of consensus"
4092..4231
/note="L1M4 repeat: matches 612..756 of consensus"
4373..4559
/note="LMC repeat: matches 971..1152 of consensus"
4586..4619
/note="17 copies 2 mer aa 100% conserved"
4623..4964
/note="MLT1A1 repeat: matches 3..333 of consensus"
4989..5149
/note="L1M8 repeat: matches 6120..6288 of consensus"
5152..5602
/note="L1PA15 repeat: matches 5325..5779 of consensus"
5606..5916
/note="Alusg repeat: matches 1..313 of consensus"
5917..6014
/note="49 copies 2 mer ga 75% conserved"
6018..6409
/note="L1PA15 repeat: matches 5769..6157 of consensus"
6566..6750
/note="L1M4 repeat: matches 2857..3036 of consensus"
6751..7063
/note="Aluud repeat: matches 1..307 of consensus"
7064..8097
/note="L1M4 repeat: matches 3036..4105 of consensus"
8099..8419
/note="Alusx repeat: matches 1..306 of consensus"
8420..8487
/note="LMC/D repeat: matches 5582..5648 of consensus"
10039..10396
/note="THE1B repeat: matches 1..364 of consensus"
10519..10807
/note="Alusg repeat: matches 3..297 of consensus"
11343..11416
```

```
repeat_region
11448..11983
/note="37 copies 2 mer tt 67% conserved"
repeat_region
12030..12337
/note="L1MD3 repeat: matches 7162..7739 of consensus"
repeat_region
12995..13077
/note="Alusx repeat: matches 1..309 of consensus"
repeat_region
13110..13222
/note="MIR repeat: matches 103..186 of consensus"
repeat_region
13228..13518
/note="FLM.C repeat: matches 2..124 of consensus"
repeat_region
13519..13853
/note="LMC repeat: matches 2131..2414 of consensus"
repeat_region
13854..14039
/note="MERA repeat: matches 1..346 of consensus"
repeat_region
14040..14254
/note="LMC repeat: matches 1951..2131 of consensus"
repeat_region
14255..14807
/note="MER20 repeat: matches 1..218 of consensus"
repeat_region
14808..15105
/note="LMC repeat: matches 1686..1951 of consensus"
repeat_region
15106..15492
/note="Alub repeat: matches 1..311 of consensus"
repeat_region
15496..15822
/note="LMC repeat: matches 1309..1686 of consensus"
repeat_region
15824..15920
/note="L1M5 repeat: matches 5823..6167 of consensus"
repeat_region
15885..16086
/note="LMC repeat: matches 1044..1140 of consensus"
repeat_region
15920..16082
/note="match: STS: Em:G34444 Em:L14255"
repeat_region
16067..16294
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
16923..16998
/note="match: GSS: Em:A0769824"
repeat_region
17107..17166
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
17175..17755
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
17839..18106
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
18213..18318
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
18706..18811
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
18812..19057
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
19058..19492
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
19556..19872
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
19873..21594
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
21587..22280
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
22298..22742
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
22743..23048
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
23049..23286
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
23285..23361
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
23364..23429
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
23429..24305
/note="match: GSS: Em:A054794 Em:A0823559"
repeat_region
24306..24593
/note="match: GSS: Em:A054794 Em:A0823559"
```


ACCESSION	AF208161
VERSION	AF208161.1
KEYWORDS	GI:6760400
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2930)
AUTHORS	M,S.,Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L., Lavallee,E., Tang,X.Y., Edouard,P., Howes,S., Keith,J.C. Jr. and McCooy,J.M.
TITLE	Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis
JOURNAL MEDLINE REFERENCE	Nature 403 (6771), 785-789 (2000) 2 (bases 1 to 2930)
AUTHORS	Sha,M., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L., Lavallee,E., Tang,X., Edouard,P., Howes,S., Keith,J.C. Jr. and McCooy,J.M.
TITLE	Direct Submission
JOURNAL FEATURES	Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park Drive, Cambridge, MA 02140, USA Location/Qualifiers 1..2930
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="AJ172-2s" /tissue_type="testis" 1..2930 /tpc_family="Human endogenous retrovirus HERV-W"
repeat_region	1..929 930..2546 /note="envelope protein" /codon_start=1 /product="syncytin precursor" /protein_id="AAF28334.1" /db_xref="GI:6760401"
5'UTR CDS	/translation="MALPYHLFFLTLLPFLTLAPPCCRMKSSPFYQELFMMRRPP GNDAPSYRLSKGTPTFTATTHPRNCYSATLCMANHTMYGMKINSCPGGLGV TWCWTFYTQMSGGVGVDARKEHYKSOLTRHGSSPKGDLSKHETLRT HTRVSLFNITTLTGILHEVASONPNMCICLPNRPVASTIPPEOMNNPTEINTTSV LVGLVSNIETLTHSNLTCKVFSMTTYTTNSOCIRWTTPROYICDPSGIFPVCGTA YRCNGSSEKCFSLFVPPTITTEDDLNYVISFRNKRPILFPVICAGVLGAIG TGIGITTSTQFYFKLSOELNGDMERADSLVTLODJNLISLAAYVIONRALDLLAE RGSGCTFGEECCCYVNQSGIVYEKVIEIDRIORAELELRNTGPMDLSOWEPIIP FIAGLIAILLILEGPCIFNLNVFVSRSRIEAVALQNEPMKOSKTKIYRRPLDRPASP RSDVNDIKGNPPEETISAQPILRRNSAGGS" 930..989 990..2543 /product="syncytin" 1869..1880 /note="furin cleavage site" /evidence=not_experimental 2043..2120 /note="Region: immunosuppressive region" 2235..2337 /note="transmembrane-region site" 2547..2930
misc_feature	3'UTR
BASE COUNT	842 a 800 c 571 g 717 t
ORIGIN	
Query Match	90.2%, Score 1335.4; DB 9; Length 2930;
Best Local Similarity	93.9%; Pred. No. 0;
Matches 1390: Conservative	0; Mismatches 91; Indels 0; Gaps 0;
1	atggccctcccattacaccttcctcttaactggtctcttaaccctcgctctaact 60
Dbb	atggccctcccattacacctttttctttcttactggtcttttttaccctctttactcttcaact 989
Oy	gcaccctcccatagtctgltgaacaaccagtagctcccccttacccaagaagtttatgaaga 120
Dbb	gcaccctcccataggcggctgrrtgacacgatgacctccctttgccagaagatttttatggaga 1049

OY	121	acggagcttcctggaatatgtgtgcgccatacataaggaagttaataagggaacacc	180
Db	1050	ATGAGAGGTCCGGAAATATTTGATGCCCCACTGTAAGAGCTTCTPAAGGAACCCCC	1109
OY	181	acctcaactgcccacaccacataatgcccgaaatgtgataactctgcacactcttgca	240
Db	1110	ACCTTCACTGCCCCACACCCTATGCCCCGCACTGCTACATCTCTGCCACTTTGCATG	1169
OY	241	catgcataatactcaatatttgacagggaaaaatgatbaatctcagttgtctcttgaggact	300
Db	1170	CATGCAAAATACTCTATTGTGGACAGGAAAAATGATTAATCTAGTTGCTCGAAGACTT	1229
OY	301	ggagcaactgtctgtgtgacttaacttcacccataccagatgtctcgaatgggggtgaaat	360
Db	1230	GGAGTCACTGTCTGTGGACTTACTTCCACCAACGATATCTGTGATGGGGGTGAACTT	1289
OY	361	caagtcagggagagaagaanaacaagttaaggaaagcaatctcccaactgacccggggaaat	420
Db	1290	CAGATCAGGCGAAGAAAAACATGTAAAAGAAGTATTTCCCAACTCACCCGGGTACAT	1349
OY	421	agcacccctagccccctacaaaggactagtctctcacaacctacatgaaacctccgttacc	480
Db	1350	GGCACTCTAGCCCCCTACAAAGAGCTATGCTCTCAAACTACATGAAGAACCTCGGTACC	1409
OY	481	catactgcctgtgtgagcctatttaatacaccctcaactcgcgtccatggaagtctcagc	540
Db	1410	CATCTGCCCGGTGTAACCTCAATTTAATACCACTCTACGGGCTCCATGAGGTCTCGACC	1469
OY	541	caaaacccctactaaactgttgatgtgcctcccccgcactcttaagcgbatacttcaatc	600
Db	1470	CAAAACCTACTAATCTTTGGATATGCTCTCCCTTAACCTTAGGCCATATTTTCAATC	1529
OY	601	ccgtctccatgaacaatggaacaactctcagcacagaataaacaacactccgttttagta	660
Db	1530	CCTGACTCGTAACATGAGAACCACTTCAGACAGAAATTAACACCACTCCCTTTTACTA	1589
OY	661	ggacctctgttttccaatctggaataaacccataccatactcctaactcgtgtaaattt	720
Db	1590	GGACCTCTGTTTCAATCTGGAAATTAACCCATACCTCAACCTCACCTGTGTAAATTT	1649
OY	721	agcaatactatagaacaacaacagctcccaatgcatcaggtgggtataacacctcccaaga	780
Db	1650	AGCAATATCTACATACACAAACCACTCCCAATGATAGTGGGTAACTCTTCCACACAA	1709
OY	781	atagctctgctcaccctcaggaaatatcttctctgtgtgtaactcagcctatcatgtttg	840
Db	1710	ATAGCTGTCCTTAACCCCTCAGGAATATTTTTGTGTGTGACCTCAGACCTATGTTGTTG	1769
OY	841	aatgctcttctaagaatctatgtgctctctctctcatctttagtgccccctatgaacctac	900
Db	1770	AATGGCTCTTCAGAAATGATGTGCTTCCTCTCAATTTTGATGCCCCCTATGACACATAC	1829
OY	901	actgacaagaattatatacaatctgtctgtaaccagccccacaacaagaatgacctat	960
Db	1830	ACTGACAAAGTTTATATCAATTATGTCTATTTCTAAACCCCGACAAAAGAGTACCCTT	1889
OY	961	cttccctttgttalcagaagcagaggatgtctaggcagactagtiactgtgcatcttgccagtatc	1020
Db	1890	CTTCCCTTTTGTATAGGAGCGAGGAGTGCTGTGATGACCTAGTACATGCGATTGGCGTATC	1949
OY	1021	acaacctctactcagttctactatacaactctccaagaataaagtgtgacatggaacag	1080
Db	1950	ACAACCTCTACTCGTTCTACTACAAACATATCTCAAGAACTAAATGGGACATGGAAAGG	2009
OY	1081	gtcactgactccctcgtgcactcttgcaaatcaaatcttaactccctagcagctagtctct	1140
Db	2010	GTCGGCAGACTCCCTGGTCTACCTTGCAAGATCAACCTTAATCTCCCTAGCACACATATCTCTT	2069
OY	1141	caaaatcgaaagactttaagacttgctaaacgcccaaaagaaggggaaacctgtttatttta	1200
Db	2070	CAAAATCGAAGAGCTTTAGACTTGCTTAACCGCTGAAGAGGAGGGGAAACCTGTTATTTTAA	2129


```

RESULT 10
AX007999
LOCUS AX007999 2055 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 22 from Patent W09967395.
ACCESSION AX007999
VERSION AX007999.1 GI:9995696
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2055)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
ENDGENE retroviral motifs, and their uses
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
SOURCE
1..2055
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..87
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07392.1"
/db_xref="GI:9995697"
/translation="PKTANLVADITSLAKYQVEKTLQGTYP"
BASE COUNT 576 a 574 C 376 G 529 T
ORIGIN

```

Query Match	90.1%	Score 133.8	DB 6	Length 2055
Best Local Similarity	93.8%	Pred. No. 0		
Matches 189	Conservative	0	Mismatches 92	Indels 0
QY 1	atggccctccatcatcattctctcttactgtgtcttaaccccttgctcact	60		
DB 391	atggccctccatcatcattctctcttactgtgtcttaaccccttgctcact	450		
QY 61	gcaccctccatcgtcgtcgtcaaacagtagctcccttacaagagttctatgaaga	120		
DB 451	gcaccctccatcgtcgtcgtcaaacagtagctcccttacaagagttctatgaaga	510		
QY 121	acggcgttctctcgtggaatatgtatgcccatatagatgatttactaaggaactcc	180		
DB 511	atggcgcgtcccggaattttgatgcccatggtatgaagagctttctaaagggaaccccc	570		
QY 181	accttaactgcccacacacccatatgcgccgcaactgctataactctgcactcttcgacg	240		
DB 571	accttaactgcccacacacccatatgcgccgcaactgctataactctgcactcttcgacg	630		
QY 241	catgcaaatcattcatatttgaacagggaataatgatbaatcactagttgttccttgaagact	300		
DB 631	catgcaaatcattcatatttgaacagggaataatgatbaatcactagttgttccttgaagact	690		
QY 301	ggaagcaactgtctgttgaacttactaccacacacagatagtctcgtatggagtgtaatt	360		
DB 691	ggaagcaactgtctgttgaacttactaccacacacagatagtctcgtatggagtgtaatt	750		
QY 361	caaggtcagaagagaagaaaaacaagttaaaggaagcaattcccaactggcccggaact	420		
DB 751	caaggtcagaagagaagaaaaacaagttaaaggaagcaattcccaactggcccggaact	810		
QY 421	agcaccctcagccctcacaagaagactagttctctcaaaactcaatpaaacctccgtacc	480		
DB 811	agcaccctcagccctcacaagaagactagttctctcaaaactcaatpaaacctccgtacc	870		
QY 481	catactgccttggttagcctatttaataccacccttaactcgtgctcctatgaagtctcagcc	540		
DB 871	catactgccttggttagcctatttaataccacccttaactcgtgctcctatgaagtctcagcc	930		

QY	541	caaaacccctactaactglttgatgctgcctcccccgcaccttcagggcaatacattcaatc	600
Db	931	CAAAAACCCCTACTACTGTTGGATTATGCGCTCCCCCGAACTTTCAGGCCATATGTTTCAATTC	990
QY	601	ccgtcttcacgaacatggaacaacttcagcaacagaaataaacaacattccgttttagta	660
Db	991	CCTGTACTCTGGAACAATGGAGACAATCTTCAGCACAGAAATTAACACACACTTCCGTTTTAGTA	1050
QY	661	ggacctctgtttccaatctggaaataaccccaataccctaaacctcaacctgtgtaaaatt	720
Db	1051	GGACCCCTTGTTTCCAAATCTGGAAATTAACCCATACCTCAAAACCTCAACCGTGTTAAATTTT	1110
QY	721	agcaatactatagacacaacaacagctcccgaagatcagatggtggtgtaaacctccacaga	780
Db	1111	ACCAATACTACTACTATACACAACCAACCTCCCAATGTCATCAGATGGGTGTAATCTCCCAACAA	1170
QY	781	ataagctgcctacccctcaggaatatctttgtcgtgtgtaacctgaacctatcattggttg	840
Db	1171	ATATGCTCGCCTTACCTCAGGAATATTTTTTGTCTGTGTGTTACTTCAGCCTATGCTTGTGTTG	1230
QY	841	aatggctcttcagaaactatgctgcttcctcattctcttagtgcacctatagcaactac	900
Db	1231	AATGGCTCTTGAGAAATCTATGTGCTCCGCTCATTTCTTAGTGCCCCATATGACATCTAC	1290
QY	901	actgacaagaattatatacaatcatgtcgtacctaaagccccacacaaaagataccat	960
Db	1291	ACTGAAACAAGATTTTATACAGTTATGTCAATCTTAAGCCCCCGCAACAAAAGATACCCATTT	1350
QY	961	cttcctcttgatatacagaacagagtgctcaggcagactcaggtatgcctgagatggcaatc	1020
Db	1351	CTTCCCTTGTGTATAGAGACACGAGATGCTTAGTGTCATGATGACTGATGGCATTTGGCCGTATTC	1410
QY	1021	acaacctctactcaagttctcactcacaaactatctcaagaataataatgtygtacatggaacag	1080
Db	1411	ACAACCTCTACTCAGTTTACTTACTACAACATATCTCAAGAACTTAATGGGACATGGAACGG	1470
QY	1081	gtcaactgaactccctggtcactccttcgaagatacaacttaactccctcagcagtagtccct	1140
Db	1471	GTTCGCGCATCTCCCTGGTCACTTGGCAAGTCAACTTAATCTCCCTGACGACATATCTCTT	1530
QY	1141	caaaatcgaagagccttagaactgtctaaacgcacaaagaggggagaaacctggtatttta	1200
Db	1531	CAAAATCGAAGACCTTTTAGACTTGTCTAACCCGTGAAGAGGGGGAACTCGTTTATTTTTHA	1590
QY	1201	ggagaagaacgctgttbatatgttcaatcaatccagaaatgttcaactgagaaagttaaagaa	1260
Db	1591	GGGGAAGAATGCTGTTATATATGTATATCAATCCGGAATGTCACAGAGAAATTAAGAA	1650
QY	1261	attcgagatcgaaatacaatgtatagagcagagaagctctcaaaacacaggaacgctggygcctc	1320
Db	1651	ATTTCGAGATCGAATATCAACGTTAGACAGAGAGAGCTTCGAAACATCTGGAGCCCTGGGGCCTTC	1710
QY	1321	ctcagacgaatgagatgcacctgggtttccacctctcttaggacccctcagcagccttaatatg	1380
Db	1711	CTCAGCGAATGGAATGCCCTCGGATTTCTCCCTTCTTAGGACCTCTTAGCAGCTTAAATATTG	1770
QY	1381	ttactctctcttggaacctgttactttaacctctgtttaaagtgtgtctctccagaatt	1440
Db	1771	CTACTCTCTCTTGGAACCCGTATCTTTAACTTCCTTTGTTAACTTTGTCTTCCAGAAATTC	1830
QY	1441	gaagctgtataagctacagatggtcttcaaatatgaaaccca	1481
Db	1831	GAAAGCTGTAAACTCAAAATGAGACCCCAAGATGACAGTCCAA	1871

RESULT	11			
LOCUS	AX007978	2559 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO967395.			
ACCESSION	AX007978			
VERSION	AX007978.1	GI:9995675		
KEYWORDS	.			


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="H_RG083M05"
/cclone_lib="CIB-978Sk-B"
/complement(838..1131)
/rpt_family="ALU"
<1360..16971
/gene="WUGSC:H_RG083M05.1"
join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"
/feature="WUGSC: strong similarity to peroxisome
biosynthesis protein PAS1 (PID:g1172019); coded for by
human cDNA C04279 (NID:g1467530)"
/codon_start=1
/protein_id="AAB46346.1"
/db_xref="GI:1669371"
/translation="KRLNIOKTLFVAFSEAVWMOQSVYLLDDDLIAGLPANPEHEH
SPDAERCEILCNVIRKNDICDKINFTDLDQHVAKETGCFARDDTVLVDRATISRL
SHOSISTREKLVTTLDLFOKALRGFLPASLRVNLHPRDLGMDKIGLHEVQKQIMD
T1QLPAKVCLEKKEYPDLNPLIRQRTGILLYPGTGKTLGAGTJARESRMNTISV
KGBELLSKYGASEQAVRDIETRAQAKPCILFDEDESIAPRGHNTGVDTRVVO
LTOLDGVEGLOGVYVAATSRPDLIDPALRGRLDKCYGCPPOVITSYLESKTO
OMLHSFLVRLTELNVLSDSLPLADVDIOHVASVYDSFGADLKALVYNAOLEALHG
MLSKSEILPDESKRMKRLYFGSSYESELNGTSSDLSQCLSNPSSMTQDLPEVP
KRDQLFSPPVLTATASEGCELTQERDQLRADISIIKGRVRSQSGEDSMQPPPI
KTRLAISOHLMTALGHTRPISSEDDMKNFAL"
/complement(4948..5130)
/rpt_family="ALU"
complement(6581..7133)
/rpt_family="L1"
complement(7767..8037)
/rpt_family="ALU"
complement(8186..8472)
/rpt_family="ALU"
8473..8625
/gene="WUGSC:H_RG083M05.1"
/feature="WUGSC: match to human 3' EST H75782 (NID:g1049794), bases
287-444"
8841..9161
/gene="WUGSC:H_RG083M05.1"
/feature="WUGSC: match to human 5' EST H75921 (NID:g1050050), bases
21-348"
9481..9547
/gene="WUGSC:H_RG083M05.1"
/feature="WUGSC: match to human 5' EST H41382 (NID:g117434), bases
143-266"
13794..13877
/rpt_family="ALU"
13878..13906
/gene="WUGSC:H_RG083M05.1"
/feature="WUGSC: match to human 5' EST H41382 (NID:g117434), bases
30-58"
13907..14104
/rpt_family="ALU"
complement(14110..14137)
/rpt_family="L1"
complement(15618..15907)
/rpt_family="ALU"
17327..17522
/rpt_family="ALU"
18667..19235
/feature="WUGSC: match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476

```

```

(NID:g759399)"
19550..19670
/rpt_family="ALU"
21507..37303
/feature="WUGSC: similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:g842257)"
37316..37489
/feature="WUGSC: Grail prediction, score = 80"
/feature="WUGSC: not experimental
complement(38938..39224)
/rpt_family="ALU"
39225..39707
/feature="WUGSC: match to multiple human ESTs, see N30113
(NID:g1148633)"
39800..40085
/rpt_family="ALU"
complement(40247..40538)
/rpt_family="ALU"
complement(40632..40924)
/rpt_family="ALU"
complement(42283..42891)
/rpt_family="ALU"
complement(45474..45613)
/rpt_family="ALU"
complement(45614..45737)
/feature="WUGSC: match to human 3' EST H48898 (NID:g988738), bases
129-333"
complement(46107..47026)
/feature="WUGSC: match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and M78831
(NID:g273146)"
complement(47027..47318)
/rpt_family="ALU"
complement(47365..47782)
/feature="WUGSC: match to multiple human ESTs, see W37495
(NID:g1319089)"
47898..48115
/feature="WUGSC: match to human 5' EST H62306 (NID:g1015138), bases
93-368"
complement(48116..48405)
/rpt_family="ALU"
complement(48406..48584)
/feature="WUGSC: match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"
complement(48787..49405)
/rpt_family="ALU"
complement(49406..49534)
/feature="WUGSC: match to human 3' EST R65794 (NID:g838432), bases
309-440"
complement(49638..49672)
/rpt_family="ALU"
complement(49674..49890)
/feature="WUGSC: match to human 3' EST N29952 (NID:g1148472) and 5'
EST N29938 (NID:g1148458), sequences are from opposite
ends of the same clone"
complement(49698..51806)
/gene="WUGSC:H_RG083M05.2"
complement(join(49698..49888,51575..51806))
/gene="WUGSC:H_RG083M05.2"
/feature="WUGSC: coded for by human cDNAs M37389 (NID:g1319205),
R65891 (NID:g838529), R65794 (NID:g938432) and R65794
(NID:g838432)"
/codon_start=1
/protein_id="AAB46345.1"
/db_xref="GI:1669370"
/translation="MFYFYQCGIIFCFPGVYVQIGDVSVYDEODGKPYAQINGF
IDOVCEKSAITMLPIPTSSPRDQDPASVYIIGPEEDLPKMEYLEFVCHAPSEYFK
SRSSPPYVTPREKGYIWHVGPPTAITIKESVANHL"
complement(51576..51758)
/gene="WUGSC:H_RG083M05.2"
/feature="WUGSC: Grail prediction, score = 86"
/feature="WUGSC: not experimental

```


repeat_region complement(52052..52329)
/rpl_family="L1"
misc_feature 55557..55843
/note="match to human EST M79192 (NID:q273505) base 2-289"

Query Match 90.1%; Score 1333.8; DB 9; Length 56093;

Best Local Similarity 93.8%; Pred. No. 0;
Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

1 atggccctccctatcatcatcttctccttactgtctccttaccctctgctcact 60
35879 ATGGCCCTCCCTATCATATTTTCTCTTACTGTTCTTTTACCTCTTTCACCTCACT 35938
61 gcaacccctccatctgtctgtacaaacagtagctcccttaccagagttctatgaaga 120
35939 GCACCCCTCCATGCGCGGTATGACAGTACTCCCTTACCAAGAGTTCTATGAGAGA 35998
121 acggcgctccctgggaatatgtatgcccacatcatatagagtttcttaaggaaaccc 180
35999 ATGCAAGCGTCCGGAAATATATGATGCCCATGTATAGAGTCTTCTTAAGGGAAACCC 36058
181 accctacatgcccacacacacatactgcccgaactgtatatactctgcccactcttgcac 240
36059 ACCTTACATGCCACACACCATATGCCCGCACACTGCTATCACTCTGCCCTTTGCAATG 36118
241 catgaatatcatcatatgtgacaggaagaatgatataatcctagtctgtctggaagact 300
36119 CATGCAATATCATATGATGATGACAGAAATGATTAATCTAGTTGTCCTGAGAGACTT 36178
301 ggaagcacatgtctgtgactacttaccacacacacagatgtctgagggagtggaact 360
36179 GGAAGTACTGTGTGTGGCTTACTTCAACCACTGATGTCTGATGGGGGTGAGATT 36238
361 caaggtcagcagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 420
36239 CAAAGTCAAGCAG 36298
421 agcacccttagccctcacaagaagactagtctcacaacacatcatgaaacccctcgtacc 480
36299 GGCACCTTAGCCCTACAAAGAGACTAGATCTCAAAATCATAGAAACCTCCGTATCC 36358
481 catctgcctgtgtgagcctatttaacacacacacacacacacacacacacacacac 540
36359 CATCTGCGCTGTGATGCTATTATATACACCTGCTGAGGCTCATGAGGTCTGCGGC 36418
541 caaaacccctactactgtgtgagtgtcctccctcgtcactctcagggccatatactcaac 600
36419 CAAAACCCCTACTACTGTTGATGATGCCCTCCCGCTGAACCTTCAAGCCATATTTCAATC 36478
601 cctgtctcgaacaatgaaacaactcagacagagaataaacaacacactcgttttaagta 660
36479 CCTGTACCTGGAACATGGAACACTTCAGACAGAAATMAACACCACTCCCTTTTAGTA 36538
661 ggaactctgtttcacaatgtgaaataacacacacacacacacacacacacacacacac 720
36539 GGAACCTGTGTTCAATGTGAAATTAACCACTCAACCTCAACCTCACTGTATTAATTT 36598
721 agcaatatcatagacaacaacagcctcccatgcatcaggtgggtaaacctccacaaga 780
36599 AGCAATATCATATACAAACCAACCTCCCAATGCATCAGTGGTAACTCTCCACACAGA 36658
781 atagctgcctacacccctcaggaatattttgtctgtgtaacctcagccatcatgtttg 840
36659 ATAGCTGCTACCTACCTCAGGATATTTTGTGTGTGATCACTCAGCCATATGTTGTTTG 36718
841 aatgctctcagaatcatatgtgtctcctcctcatctctagtgccccctatgacatcac 900
36719 AATGCTCTCTTGAGAAATCATATGTGCTTCTCTATTTAGTCCCTCATATGACATCTAC 36778
901 actgaacaagaattatatacatgtctgtaacagccccacaacaagaatataccact 960
36779 ACTGAACAAGATTTATACAGTTATGTCAATATCTAAGCCCCCGCAACAAGAGATACCAATT 36838

961 ctctcttgtatataagcagagagtgctaggcagactagtgatctgcaatctgcaatctc 1020
36839 CTTCCTTTTGTATAGAGCAGAGAGTGTAGTGCATGCTAGTACGATTTGGCGGTATC 36898
1021 acaactctaccatgtctctacacaaactatctcaagaataaagtgagacatggaagaag 1080
36899 ACAACCTTACTCACTGTCTACTACCAACTATCTCAGAACTAAAGGGGACATGGAACGG 36958
1081 gtcaatgactccctgtgcaactctgcaagaatcaacttaactccctagcagcaatgctct 1140
36959 GTGCGGAGATCCCTCTGTGTACCTTGCAAGATCAACTTAATCTCCGTAGCAGATAGTCT 37018
1141 caaatcgaagaagctttagactgtctaacccgcaaaagaagggggaacctgttatctta 1200
37019 CAAATTCGAAGAGCTTTAGACTTCTTAACCGCTGAAGGGGGGGAACCTTTATTTTAA 37078
1201 ggaagaagaagctgttatatgtatataatcaatcagaatgtcactgagagaagttaaga 1260
37079 GGGAAAGATGCTGTATATATATATATCAATCCGAATCGTACTGAGAAATTTAAAGAA 37138
1261 atcagatcgaatacaatgttagagcagaagagacttcaaaacacacgacgctgaggctc 1320
37139 ATTGAGATCGAATACAAAGTAGAGCAGAGAGCTTCGAAACATCGAACCTCGGGGCTTC 37198
1321 ctcaagcaatgagtgacctggtgtctccctctcttagaactctcagactataatgt 1380
37199 CTCAACCAATGAGATCCCTGATTTCTCCCTCTTAGAGCCCTAGCAACTATTAATTTG 37258
1381 ttactcctcttggagccctgtatctttaaactcctgttaagtgtgtcctcagaact 1440
37259 CTACTCTCTTTGGACCTGTATCTTTAAACCTCTTTGTTAACTTTGTCTTCCAGAAATC 37318
1441 gaagctgtaaaagctacagatggtcttacaatgaaaccca 1481
37319 GAAGCTGTAAACTTCAATGAGGCCCAAGATGAGTCCAA 37359
RESULT 15
AC007566/C 149194 bp DNA linear PRI 06-FEB-2002
LOCUS Homo sapiens chromosome 7 clone CTB-1065, complete sequence.
DEFINITION AC007566
ACCESSION AC007566.2 GI:11181861
KEYWORDS Htg.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149194)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149194)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1999) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 3 (bases 1 to 149194)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 149194)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 149194)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington

Page 19

Db	85355	CGTGA	CTGAA	CAATG	GACAC	CTTCA	GACAG	AAATP	AACAC	CACTCC	GTTT	AAGA	85466
Db	85355	CGTGA	CTGAA	CAATG	GACAC	CTTCA	GACAG	AAATP	AACAC	CACTCC	GTTT	AAGA	85466
Qy	661	ggacctctgtgttccaatc	tggaataaccca	tactcca	aacctt	acgtgt	taaat	att	720				
Db	85465	ggacctctgtgttccaatc	tggaataaccca	tactcca	aacctt	acgtgt	taaat	att	720				
Db	85465	ggacctctgtgttccaatc	tggaataaccca	tactcca	aacctt	acgtgt	taaat	att	720				

[illegible]

Qy	781	atagctgcctcaccctcagaataattttgtcgtgtgtaacctagccatcatgtttg	840
Db	85345	ATAGTCTCCATACCTCAGGAATATTTTGTCTGTGTAACCTCAGCCATCGTGTGTTG	85286
Qy	841	aatgagctcttcagaatcatctatgtcttcctctcatctcttagtgccctatgacatctac	900

901 actgaacaagatttatatacatcatgtgtctacctaaagccccaacaacaaagagtaccatt 960

Db 85165 CTTCCCTTTTGTATAGCAGCAGAGTGTCTAGTGCACCTAGTACTGCGATTGGCGGTATC 85106

Db 85105 ACAACCTCTACTCAGTTCTACTACAACATATCTCAGAACTAAATGSGGACATGGAAACGG 85046

```

Dh 85045 GTCCGCATCCCTGTCACCTTGCAGATCACTTAACCTCCCTAGCAGACAGTAGTCCTT 84986
QY 1141 caaatcgaagaagccttagactctgtaacccgccaagaagggggaacctgttatttta 1200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dh 84985 CAAAATCGAAGAGCTTTAGACTTGTCTAAACCGCTGAAGAAGGGGGAACCGTTATTTT 84926
QY 1201 ggaagaagaagcctgttattatgttaataatccagaatgttcaactggaagaagttaa 1260
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dh 84925 GGGGAAGAAATGCTGTTATTATGTATTAATCAATCCGGAAATCGTCACTGAGAAACTTAAAG 84866
QY 1261 attcgaagatcgataatacaatgttagagcagaagagccttcaaaaaccgaacgcctggggcc 1320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dh 84865 ATTGAGATTCGAAATCAACGTAAGAGCAGAGGAGCTTCGAAACACTGGAACCTGGGGCCTTC 84806
QY 1321 ctcaagcaaatgagtcgcccctggggttcccccctcttagagaccttagagccttaaatatg 1380
Dh 84805 CTCAGCCAAATGGATGATCCCTGGATTTCCTCCCTTCTTAGAGACCTCTGACACTTAAATATTG 84746
QY 1381 ttactccctcttggaccctgttaactttaaaccctcctgtttaaagtgttcctctccagaat 1440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dh 84745 CTACTCCCTTTGGAACCCGTATCTTTAAACCTCCTGTTAACTTTGTCTCTCCAGAAATC 84686
QY 1441 gaagctgtgaaagctacagatggtcttacaatatggaaccca 1481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dh 84685 GAAGCTGTAAAACTCAAAATGAGAGCCCAAGATGACATGCCAA 84645

Search completed: June 20, 2002, 09:52:31
Job time: 11276 sec

```

Search completed: June 20, 2002, 09:52:31
Job time: 11276 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 08:18:31 ; Search time 3276.58 Seconds
(Without alignments)
5474.444 Million cell updates/sec

Title: US-09-319-156a-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagacttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Optimal number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_estc:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593.4	44.7	771	10	BI087886 602852690
2	480.8	36.2	679	12	AG113694 Pan trogl
3	475.2	35.8	683	12	AG134524 Pan trogl
4	474	35.7	669	12	AG049953 Pan trogl
5	473	35.6	681	12	AG116971 Pan trogl
6	469.6	35.3	710	12	AG096341 Pan trogl
7	456.2	34.3	670	12	AG072852 Pan trogl
8	455	34.2	712	12	AG099717 Pan trogl
9	439.2	33.0	689	12	AG121669 Pan trogl
10	438.8	33.0	921	12	BH149565 EMTQ48TR
11	435.6	32.8	679	12	AG076758 Pan trogl
12	433.6	32.6	683	12	AG102951 Pan trogl
13	430.6	32.4	611	12	AG038171 Pan trogl
14	429.4	32.3	736	12	AG086046 Pan trogl
15	427.6	32.2	714	12	AG0238712 RPII1-64
16	424.4	31.9	697	12	AG108186 Pan trogl
17	423.8	31.9	675	12	AG117095 Pan trogl

C 18	419.4	31.6	672	12	AG040241 Pan trogl
C 19	416.2	31.3	692	12	AG114407 Pan trogl
C 20	415.6	31.3	494	9	AA781423
C 21	414.6	31.2	636	12	AG087582 Pan trogl
C 22	414.4	31.2	522	9	AW971553
C 23	411	30.9	653	12	AG033781 Pan trogl
C 24	402.4	30.3	685	12	AG184837 Pan trogl
C 25	399.6	30.1	695	12	AG093463 Pan trogl
C 26	397.6	29.9	688	12	AG091257 Pan trogl
C 27	395.8	29.8	647	12	AG086167 Pan trogl
C 28	393	29.6	490	9	A1598135
C 29	393	29.6	660	12	AG047615 Pan trogl
C 30	390.6	29.4	701	12	AG126669 Pan trogl
C 31	388.6	29.2	718	12	AG118424 Pan trogl
C 32	386.2	29.1	609	12	AG066901 Pan trogl
C 33	383.6	28.9	440	10	BE732673
C 34	383	28.8	687	12	AG178479 Pan trogl
C 35	382.6	28.8	706	12	AG091989 Pan trogl
C 36	381.8	28.7	644	12	AG068897 Pan trogl
C 37	381.4	28.7	662	12	AG088476 Pan trogl
C 38	381.2	28.7	543	10	BI963185
C 39	380	28.6	651	12	AG058970 Pan trogl
C 40	377	28.4	631	10	BG573364
C 41	376.4	28.3	648	12	AG018664 Homo sapl
C 42	376.2	28.3	641	12	AG036829 Pan trogl
C 43	374.4	28.2	661	12	AG049502 Pan trogl
C 44	372.8	28.1	485	9	AW511366
C 45	370.4	27.9	687	12	AG118071 Pan trogl

ALIGNMENTS

RESULT 1
BI087886 771 bp mRNA linear EST 20-JUN-2001
LOCUS 602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
DEFINITION mRNA sequence.

ACCESSION BI087886
VERSION BI087886
KEYWORDS GI:14506216
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

NIH-MGC http://mgc.ncl.nih.gov/
1 (bases 1 to 771)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
Tissue-Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LAM11015 row: d column: 23

High quality sequence stop: 762.
Location/Qualifiers

FEATURES

source

1..771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4993894"
/clone_id="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 225 a 206 c 166 g 174 t

ORIGIN

Query Match 44.7% Score 593.4: DB 10: Length 771;
 Best Local Similarity 89.2% Pred. No. 2e-159;
 Matches 688: Conservative 0; Mismatches 66; Indels 15; Gaps 4;

QY 107 gagaagatgaagaatctgagatcgaatataatgtagagagagagacccaact 166
 DB 1 GAGAAAGTAAAGAAATTCGATGCAATATACAAAGAGAGAGAGCTTCGAAACACT 60
 QY 167 gacacctgggggctctctcaagcagatgagatgacctctccctcttaagacctca 226
 DB 61 GGACCCCTGGGGCTCTCTCAGCAATGAGATGCCCTGATTCCTCCCTCTTACGACCTCA 120
 QY 227 ggaagctataatcttctctctctctctctctctctctctctctctctctctct 286
 DB 121 GGAGCTATATATTTGCTACTCTCTTTGGACCTGATCTTAACTCTCTTTAACTTT 180
 QY 287 gctctctcagaatctgaagctgtlaaagctacaatagttctcaaatggaaccagatg 346
 DB 181 GTCTCTTCAGAACTGCAAGCTGTAATACTA-----CAATGAGGCCCAAGATG 228
 QY 347 cagtcctgactcaaaatctaacctgtgagacccctggagacccctgctagactatgctgat 406
 DB 229 CAGTCCAAACTAAAGATCTACCGCAGACCCCTGGACCCGCTGTAAGCCAGATCTGAT 288
 QY 407 gttatgacattggaagtcacccctcccgaggaatctcaactgcaaacccctactacac 466
 DB 289 GTTATGACATCAAAAGCCACCCCTCTGAGAAATCTCAGCTGCACACACTTACTACGC 348
 QY 467 tccaattcagtagaagcagtagagacagttgtcagaccaactccccaagtagtctggg 526
 DB 349 CCCAATTTCAGCAGGAAGCTTAAAGCGGTGCTC-GGCNACTCCCAAAACACACTTAG 407
 QY 527 tttctctgtg-tgaagggtgtgagctgagagacagagtagctgagattcttaagctgagca 585
 DB 408 TTTCTCTGTGATGAGTGGGGAGTACAGACAGACTACTGATTTCTTACGGCTGACCTA 467
 QY 586 agaattcccaagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 645
 DB 468 ACAATTCCTTAACCTTAAGCTGAGAGTGAACATTCACACTTAAACAGGGGCTTGCA 527
 QY 646 cttagctcaccgacccaatcagagagctcactaaatgcttaactgaagcaaaacagga 705
 DB 528 CTTACTCTCACCTGACCAATCAGAGAGCTCACTAAATGCTAATTAAGCAAGACAGA 587
 QY 706 gcta-aagaaatagaacatcatctatgtctgaagagacagggaggaagagagattgg 764
 DB 588 GTTACAGAAATAGCCAAATCATCTATGCTGAGAGCACAGAGGAGGACATGATCG 647
 QY 765 gataaaatcagagcatcctcaagcagcaaacacccctctgggtccctccatgtc 824
 DB 648 GATATTAAACCAAGCTTGAGCCGCGCAACGGCAACCCCTTTGGGTCCCTCTTGT 707
 QY 825 atggagagctcgtcttctactatctactatcactatcaaatcagcaactgca 875
 DB 708 ATGGGAGCTCTGTATCATGCTATTTCACTTAAATCTTGAACACTGCA 758

RESULT 2
 AG113694/c AG113694 679 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-120G11.R, genomic survey sequence.
 DEFINITION AG113694
 ACCESSION AG113694
 VERSION AG113694.1 GI:16734213
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (sites)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokok, Y., Watanabe, H., and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 679)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokok, Y., Watanabe, H., and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimpesgsc.riken.go.jp, URL:http://nmp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.
 COMMENT PRIMERS
 Sequencing: M13rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 FEATURES
 source location/Qualifiers
 1. 679
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-120G11.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"
 BASE COUNT 168 a 177 c 189 g 144 t 1 others

ORIGIN

Query Match 36.2% Score 480.8: DB 12: Length 679;
 Best Local Similarity 86.7% Pred. No. 4.5e-127;
 Matches 552: Conservative 0; Mismatches 83; Indels 2; Gaps 2;

QY 673 gctcaataaattgctcaatcaggaacaaacagagaggaagcaatagccaatcatctatg 732
 DB 678 GCTCGTAAAGATGTAATTAAGCAAAACAGAGGTAAGAAATAGCCAAATCATCTATG 619
 QY 733 cctgaagacagcgagggaagcaagagatggtgataataactcagagcatcaagccaaga 792
 DB 618 CGTGAGAGCACAGTGTAAAGAAAGGGTCGGGATTAAGCCCGGCAATTCAGCGCGCA 559
 QY 793 ac-agcaaacccctctgggtccctcccaatgtagggagctc-gtttcaactatct 850
 DB 558 ACGGCGAGCCCTCTTGGGTCCCTCCCTCTTGTATGACAGCTCTGTTTCACTATTT 499
 QY 851 cactctatataatcagcaactgcaactctctctgctcgtgtttttatagctcaagctga 910
 DB 498 CACTATATTAGATCTTGACGTGACACTCTTGTGCGGTTGTTTACAGCTCAAAACAGA 439
 QY 911 gctttgttcgcatccacacactgctgtttgcacccgtcaagaccgctgtgacttcc 970
 DB 438 GCTTTGTGCTTGGCGTCGCGCACTGCTGTTTACCGCGCTCCAGACCGCACTGACTTC 379
 QY 971 atcccttggatccagcaagatgtcgaactgtctcctgataccagagagtagcaattgcc 1030
 DB 378 ATCCCTCTGGATCCGCGAGGGTGTGCGCTGTCTACTATTCACGAGGACCCATTGCT 319
 QY 1031 actcccgatcagagctaaagagcttgccatgttctcgtcagtagctgaagtgctgtgttc 1090
 DB 318 GCTCCCGATCGGGCTAAAGGGTGTGCATGTTCTGCAATGCTAAGTGCTGCTGGTTCGTC 259
 QY 1091 ctaataagaactgaacactggttcaactggtgttcaatgttctcttcaatgaacagagcttc 1150
 DB 258 GTAATGAGACTTAACACTAGTGTGAGATTCACAGTTCTCTTCCGTGACCCACGCTTG 199
 QY 1151 taatagagctataacactcagcagtagtggccaagatctcatcctctgtgtatctggaagc 1210
 DB 198 TAATGAGACTGTACGCTACCGCATGGCCCAAGATTCCGTTCTTGGAAATCGTGAGGC 139

QY	1211	caagagccccaggtcagagaangtfgagggctgtccaccatttgggaatfgccccactgcga	1270
Db	138	CAAGAACCCCGAGGTGAGAGAACAGAGGCTTGCCACCATTCTTGGAAATGAGGCCCGCTGCACA	79
OY	1271	tttctgttagcgccaccacacatcttggagactgtg	1307
Db	78	TTTTGGAAGTGGCCCGCCACCATCTTGGAGACTCTCG	42
RESULT	3		
LOCUS	AG134524	683 bp	DNA
DEFINITION	Pan troglodytes DNA, clone: PTB-147110.R, genomic		GSS 04-NOV-2001
ACCESSION	AG134524		linear
VERSION	AG134524.1	GI:16664202	genomic survey sequence.
KEYWORDS	GSS: GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
TITLE	1 (sites)		
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,		
REFERENCE	Totohi, Y., Watanabe, H. and Sakaki, Y.		
AUTHORS	BAC end sequences of library PTB		
TITLE	2 (bases 1 to 683)		
JOURNAL	Unpublished		
COMMENT	Totohi, Y., Watanabe, H. and Sakaki, Y.		
	Direct Submission		
	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical		
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);		
	1-7-22 Suehichi-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan		
	(E-mail: chimbes@sc.riken.go.jp, URL: http://bgp.gsc.riken.go.jp/		
	Tel:81-45-503-9111, Fax:81-45-503-9170)		
	Clones are derived from the chimpanzee BAC library PTB This BAC end		
	was generated during the R&D process and may have higher chance of		
	clone tracking errors.		
	PRIMERS		
	Sequencing: MJ3Rev		
	LIBRARY		
	Vector : pKS145		
	R.Site 1 : SacI		
	R.Site 2 : SacI.		
	Location/Qualifiers		
	1. 683		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
	/clone="PTB-147110.R"		
	/sex="male"		
	/cell_type="lymphoblast"		
	/clone_lib="PTB Chimpanzee Male BAC library"		
	160 a 201 c 156 g 164 t 2 others		
BASE COUNT			
ORIGIN			
Query Match	35.8%;	Score 475.2;	DB 12; Length 683;
Best Local Similarity	89.6%;	Pred. No. 1.9e-125;	
Matches 554; Conservative 0; Mismatches 60; Indels 4; Gaps 4;			
OY	666	tcaagagagctcaactaaatgtctaatacagagcaaaaacagaggtlaaagaataagccaatca	725
Db	58	TGCGNAGAGTACTTAAATGCTAATTAGGCAAAAAACAGAGGTAAAGAAATAGCCATCA	117
OY	726	tctattgcttgagagcacagcggaagacaaagattggttgatataaactcaggaatca	785
Db	118	TCTATTGGCTGAGACACAGTGGGAAGCAAGATCGGATATTAACCCAGCATCA	177
OY	786	gccagcaacagaaaccccttgggtccctcccatgtatgtgagagctgtttcaatc	845
Db	178	GCCAGCAACGGCAACCCCTTGGGT-CCATCCCTTGATGAGAGCTGTGTTTCATCTC	236

Oy	846	tatttcactctatattaacatcagaaactgcatctctctgtccggttttlttaagtctaa	905
Db	237	TATTTTCACCTCATTAATAAATTCTTGCAACTCAGTCCATCTTCTGTGCGGTGGTTGGTTGAATGCCTGCA	296
Oy	906	gcctgaactcttgctgcgatccaccacactgcgc-cttgccacacgtcaaacagaccgcctgcgtg	964
Db	297	GCTGAGCTTTCTCTCTGCCATCCACCACCTGCTGTTTTTGGCCGCCGTCGCGACGACCGCCGCTG	356
Oy	965	acttccalcaccttgtagatccagcagaagtgtccactgtgtcctctgataccagcgaggtaccc	1024
Db	357	ACTTCCATTCCTCCGAGATTCACAGCAGCGGTGTCGCTGTCTGTCOTGATCCAGGAGGCGGCC	416
Oy	1025	attgcacactcccgatcacaggtctaagaagcttgccatgtgtctctcgatagctaatgtcctggg	1084
Db	417	ATTGCCGCTCTCGAATCGGGCTTAAGAAGCTTGGCATTTGCTTCGACGCGCTAAAGTGCCTGTG	476
Oy	1085	tttgtccctaatagaactgaacaacactgctaacctggaggtgtccatgtgtctctccatgaaccac	1144
Db	477	TTTGGTCTTAATCACCTGAACACTGATGATCTACTGGTTCCACGGTTCTTCCATGATACAC	536
Oy	1145	ggctctciaatagaagcataaacaactcaaccgcatgycgccaaattccatcct-tggtatct	1203
Db	537	GACTCTTAATAGAGCTAATAACACTGACATGCGCTTAAGATTCATTTCTGCGGAATTC	596
Oy	1204	gtgagagcca-a-gaacccccagtgacagaangtgaggtctgcccacattgggaagtgc	1262
Db	597	ATGAGCGCCAATGTAAACCCACAGGTCAAGAAAAACAAGAGCTTGGCCACCACCTTGAAATGTGCC	656
Oy	1263	cactgcactttgtgtgacg	1280
Db	657	CACGACCATCTGTGGGAC	674
RESULT	4		
LOCUS	AG049953	669 bp	DNA linear GSS 02-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-030D19.F, genomic survey sequence.		
ACCESSION	AG049953		
VERSION	AG049953.1 GI:16586845		
KEYWORDS	GSS; GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,		
JOURNAL	Totoki,Y., Watanabe,H. and Sakaki,Y.		
REFERENCE	BAC end sequences of Library PTB		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 669)		
JOURNAL	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,		
REFERENCE	Totoki,Y., Watanabe,H. and Sakaki,Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical		
JOURNAL	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);		
REFERENCE	1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan		
AUTHORS	(E-mail:chimbese@sc.riken.go.jp, URL:http://ngp-gsc.riken.go.jp/,		
TITLE	Tel.:81-45-503-9111, Fax:81-45-503-9170)		
JOURNAL	Clones are derived from the chimpanzee BAC library PTB This BAC end		
REFERENCE	was generated during the Red process and may have higher chance of		
AUTHORS	clone tracking errors.		
TITLE	PRIMERS		
JOURNAL	Sequencing: -21M13		
REFERENCE	LIBRARY		
AUTHORS	Vector : pKS145		
TITLE	R.Site 1 : SacI		
JOURNAL	R.Site 2 : SacI.		
REFERENCE	Location/Qualifiers		
AUTHORS	1..669		
TITLE	/organism="Pan troglodytes"		
JOURNAL	/db_xref="taxon:9598"		
REFERENCE	/clone="PTB-030D19.F"		
AUTHORS	Source		

```

/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      151 a      185 c      160 g      173 t
ORIGIN

Query Match      35.7%; Score 474; DB 12; Length 669;
Best Local Similarity 86.8%; Pred. No. 4,1e-125;
Matches 548; Conservative 0; Mismatches 71; Indels 12; Gaps 2;

QY 669 gaaagctcctaaatgcaatcagcaaaaacagaggtaaagcaatagcaatcact 728
DB 49 GAGAGCTCACTAAATGCTGATTAGCAAAAGAGAGGTAAGAATAGCAATCATCT 108
QY 729 attgcctgagagacagcgaggaaggaaggaatctggatataaactcaagcaatcaagc 788
DB 109 ATTGCTGAGACACAGCAGGAGGAGCAATGATCGGATATTAACCCAGCATTCGACGT 168
QY 789 agcaaacagcaacccccccttggtgctccctccatctglatagagagctctgtttcactat 848
DB 169 GGCACGCGCTCTCTTGGGTGCCCTCCCTTGTATGAGAGCTCTGT-----T 218
QY 849 ttcactctatatacatatgcaatgcaatcactc--tctggtcggtgtttttatgctcaag 906
DB 219 TTCACCTATTAATCTTTCACATGCACTCTTTATGTTCAATGTTGTACAGCTCGAG 278
QY 907 ctgagcttctgtctgcacatccacacacgtgctgttgcacacgtcacagacccgtctgac 966
DB 279 GTGAGCTTTGATTTGCCGTCACACAGTGCTGTTGCTGCCGTCGACAGCCACCGCTGAC 338
QY 967 ttcacatcccttggatccagagagtgacactgtgctctglatccagcgaggtaccat 1026
DB 339 TTTCATCTCCCTCTGAGATCTGGGAGGCTGCCGCTGCTCTCATCAGAGAGTCCCAT 398
QY 1027 tgcacatcccgatcagaggttaagaggttgcacatgttccgcaatgctcaagtccctgggt 1086
DB 399 TCTCTGCTCTGATCGGATTAAGGCTTGCCATTGTTCTCGCAAGGCTAAGTCCCTGGGT 458
QY 1087 tgcctaaatagaacttgaacactgtgacactgtgctcactgtctctccatgacacag 1146
DB 459 CCTCCTAATGAGAGTGAACACTAGTCACTGAGTTCATGCTCTCTCGATCATCATAGG 518
QY 1147 ctctcaatagagctataaactcaccacgtgacccaagaatctcactctgttactgt 1206
DB 519 CTTTAAATAGACTATTAACACTCACCGCATGCGCAAGATTCATTCCTTGGAAATCCGGG 578
QY 1207 aggcacaagaaccccaaggctcagagaangtgaggtctgcacacatctggggaagtgagccact 1266
DB 579 AGGCCAAGAACCCCAAGTCAAGAAACAGAGGCTTGCCACCATCTTGGAAATGAGCACAC 638

QY 1267 gccatttggtagcgccacacacactctg 1297
DB 639 ACCATTGTGAGACAGCCCGCACCATGTTG 669

RESULT 5
AG116971
LOCUS AG116971 681 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-124K18.F, genomic survey sequence.
ACCESSION AG116971
VERSION AG116971.1 GI:16737490
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-124K18.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (bases 1 to 681)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished

```

```

REFERENCE 2 (bases 1 to 681)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Tokoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 681
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-124K18.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      155 a      207 c      154 g      165 t
ORIGIN

Query Match      35.6%; Score 473; DB 12; Length 681;
Best Local Similarity 86.8%; Pred. No. 8e-125;
Matches 538; Conservative 0; Mismatches 56; Indels 12; Gaps 2;

QY 671 gagctcactaaatgcttaatcagcaaaaacagaggttaagcaatagcaatcactat 730
DB 86 GAGCTCACTAAATGCTATTTAGGCAAAAACGAGGATTAAGCAATATACCAATCTATCTAT 145
QY 731 tgcctgaagagcagcgaggaaggaagatgagataaaactcagacatcaagccag 790
DB 146 TGCCGTGAGAGCAGACAGAGAGGACAAATGATGGATTAACCCAGCATTCGAGCCAG 205
QY 791 caacagcaaccccccttggtgtccctccatctgatatggagagctcgtttcactatctt 850
DB 206 CAACGGCTAACCTCTTGGGTCCCTCCCTTGTGAGAGCTCTGT-----TTT 255
QY 851 cactctatataatcttgcacactgtctgttgcacacgtcagacagaccgctgact 968
DB 256 CACTATTAATTAATCTTGCACATGCACTCTCTCTGCTCGTGTGTTGTTATGCTCAACT 315
QY 909 gagcttctgttcgcacatccacactgtctgttgcacacgtcagacagaccgctgact 968
DB 316 GAGCTTGTGCTTGCCATCCACACTGCTGTTTGCGCGCGCAGACACCGCCTGACTT 375
QY 969 ccattcccttggatccagagagtgatccactgtgtctctgtatccagcgagttaccatgt 1028
DB 376 CCATCCCTTCAGATCTGACAGGCTGCCACTGTGCTCTGATCCAGCAGAGCGCCACTTG 435
QY 1029 caactcccgatccataaaggcttgcacatgttctcgtcagatgaagtaagtcgtgttg 1088
DB 436 CCGCTCCCGATAGGCGCTAAAGGCTTGCCATTGTTCTCTCATAGGCTTAATGCCAGATTCA 495
QY 1089 tcttaataagactgacactgtctcactgttccatctgctctcctcctcctcagacagct 1148
DB 496 TCTTAATGAGAGCTATTAACACTCACCGCATGCGGCTCATGTGTTCTTCCGTGACCGGCT 555
QY 1149 tctaatagagctataaactcaccgcatgagcccaagatctcattcctctgtgactgtgag 1208
DB 556 TCTAATAGAGCTATTAACACTCACCGCATGCGGCAAGATTCTCTTGGAATCGTGTAG 615
QY 1209 gccaaagaccccaaggctgagagaangtgaggtcttgcacacatcttgggaagtgagccagc 1268
DB 616 GCCAAGAACCCCAAGGTCAAGAACACAGAGGCTTGCCACCATCTTGAAGTGGCCAAACAC 675

```

```
QY 1269 catltt 1274
|||||
Db 676 CATTTT 681

RESULT 6
AG096341 710 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-097K23.F, genomic survey sequence.
DEFINITION AG096341
ACCESSION AG096341.1 GI:16716658
VERSION GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-097K23.F.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 710)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
FEATURES
source
1..710
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-097K23.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 168 a 215 c 162 g 165 t
ORIGIN
Query Match 35.3%; Score 469.6; DB 12; Length 710;
Best Local Similarity 85.1%; Pred. No. 7.7e-124;
Matches 561; Conservative 0; Mismatches 65; Indels 33; Gaps 2;

QY 671 gaactcaataatgctaatacagcaaaacaggaagtaagaatgccaatcatctat 730
|||||
Db 75 GAGCTCAGTAATAATGCTAATAGCAAAAACAGAGGTAAGAAATAGCCCAATCATCAT 134

QY 731 tgcctgagagcacagcggaaggaagaaagtgtgatatataactcagcattcaagcag 790
|||||
Db 135 TGGCTGAGACACAGTGGATGACATATGACAGATATATAACCCAGCATTCAGGCTG 194

QY 791 caacagcaaccccttgggtccctccatctgtatgtgagagctctgtttcaactctatt 850
|||||
Db 195 CAACGGGAAACCCCTTGGTTCCTCTCTCTGTGTATGGAGCTCTGT-----TTT 244

QY cactcatataatcatgcaactgcaactctctctgctgctgtttttttagctcaagctga 910
|||||
Db 245 CACTCTATGCAATCTTGCAACTGCACTCTTCTGCTGTGTGTGTAGGCTCAAGCTGA 304
```

```
QY 911 gctttgtcgcaatccaccactgtctgttgcacgcgtcacagaccgcgtgacttcc 970
|||||
Db 305 GCTTTGCTCACCATCCACACACTGCTGTGGCCGCGTTGACAGCCATCGCTGACTTCC 364

QY 971 atcccttgcataccagcagagtgctcaactgtctctcgtatccagcgaggtaccattgcc 1030
|||||
Db 365 ATCCCTCCAGATCTGCGACGGGTGTCACATGCTGCTCGATCCAGCGACCATTCATGCC 424

QY 1031 accccgataagcctaagaagcttgcatctgtctcgtatggtcctaagtgctggttctgc 1090
|||||
Db 425 GCTCCCGATGGGCTTAAGGCTTGCCATGTCTCTGACGCTAAGTGCTGGGTTGCTC 484

QY 1091 ctaataagaactgaacactgtgtcactggttcacatgtgtctctccatgaccacagcttc 1150
|||||
Db 485 CTATCAAGCTGAACACTACTACTGCGTTCCATGTCTCTTCCATGACCATGCGTTTC 544

QY 1151 taatagaactataaacactcacacacatgagcccaagattccattcctgttattctgtgaagc 1210
|||||
Db 545 TAAATAGAGCTATTAACACTCACCGCATGCCCAAGATTCATTCCTTGGAATCCGTAGGCG 604

QY 1211 caagaacccaggttcagagaangttaggcttgcacacatttgggaagtggccacgtcca 1270
|||||
Db 605 CAAGAACCCTCAGGTCAAGAGAACAAGAGACT-----GCCACCA 641

QY 1271 tttagtagcgagccaccacacatcttggagctgtggaagcaagatccccaagtaaca 1329
|||||
Db 642 TCTTGAAGACGACTGCGACCATCTTGGAAGCTTAAGACCAAGACCCCCCAATAACA 700

RESULT 7
AG072852 670 bp DNA linear GSS 03-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-064H14.R, genomic survey sequence.
DEFINITION AG072852
ACCESSION AG072852.1 GI:16624654
VERSION GSS: GSS (genome survey sequence).
KEYWORDS Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
SOURCE BAC library clone:PTB-064H14.R.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 670)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenhiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
FEATURES
source
1..670
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-064H14.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC library"
BASE COUNT 152 a 190 c 157 g 170 t 1 others
```

ORIGIN

Query Match	34.3%;	Score 456.2;	DB 12;	Length 670;
Best Local Similarity	87.5%;	Pred. No. 5.4e-120;		
Matches 537;	Conservative 0;	Mismatches 64;	Indels 13;	Gaps 3;

QY	669	gaagagctactctaaatctgcatctacagcaaaacaaagagtgtaagcaatctagccaatcatct	728
Db	57	gagacctactctaaatgaagctaataggcacaacacagaggttaagaaatgaccaaactattt	116
QY	729	attgcctcgaagacacagcgaggaagagacaagagcttggatataaacctcagacgtcctaagcc	788
Db	117	attgcctcgaagacacagcagagagagacaatgatcggaatataaacctcagacgtcctaagcc	176
QY	789	agcaacagcaaacccctcttgggttccctcccatctgtataggagctctctttcactctat	848
Db	177	agcaacagcgtctacccctcttgggttccctcccatctgtataggagctctcttcttctat	226
QY	849	cttactctatctaaatcatctagcaact--gcaactctctgttcgttcttctatctgagctcaag	906
Db	227	tttactctatctaaatctggaactctggaactctgctcttcttctgctgctgtttgttactgctcaag	286
QY	907	ctgagctcttctgtcgcacatccacacactgctgttctgtccacccgttcacagaccgctgctgac	966
Db	287	ctgag--tttttgcctgctgtcgcacacactgctgttttcccccgttcacagaccctgctgac	345
QY	967	tttactctcttctgtgttcacagagagatgctgctgctgtcttctatctcagcgaggtaccat	1026
Db	346	tttactctctctcctcgcggttcacacacagaggtgctgctgtcttctatctcagcagacagccctcct	405
QY	1027	tgccactcccgatccagcgtctaaagagcttgcattgtcttctcgtcagctgaagtccttggtt	1086
Db	406	tgccactcctcgtgctgctgctaaagcgttgcattgttctcgtcagcgtgtaagtctcgtggtt	465
QY	1087	tgctctaatatgaacagcaaacactgctgctgctgctgttctcattgctcttctcattaccacag	1146
Db	466	gctctctaatcgcgctggaacactgctgctgctgctgctgttctcattgctcttctcattaccacag	525
QY	1147	cttctcaatagaagctataaacactcaccgcgatctggcccaagatctccattctctgtatctgt	1206
Db	526	cttctcaatagaagcgttataaacactcaccgcgatctggcccaagatctccattctctgtatctgt	585
QY	1207	aggccaagaacccccagctcagagaagaangtgaagcttgcacacattctgggaagtgagcccat	1266
Db	586	aggccaagaacccccagctcagagaagaangtgaagcttgcacacattctgggaagtgagcccat	645
QY	1267	gccattctgtgtagc 1280	
Db	646	gccattctgtgtagc 659	
RESULT	8		
LOCUS	AG099717	712 bp DNA linear GSS 03-NOV-2001	
DEFINITION	Pan troglodytes DNA, clone: PTB-102807.F, genomic survey sequence.		
ACCESSION	AG099717		
VERSION	AG099717.1 GI:16720234		
KEYWORDS	GSS; GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male BAC library clone:PTB-102807.F.		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryotes: Metazoa: Chordata: Gracilata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Pan.		
AUTHORS	1 (sites)		
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,		
JOURNAL	Totoki, Y., Watanabe, H. and Sakaki, Y.		
REFERENCE	BAC end sequences of Library PTB		
AUTHORS	2 (bases 1 to 712)		
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,		
AUTHORS	Totoki, Y., Watanabe, H. and Sakaki, Y.		
TITLE	Direct Submission		


```

RESULT 9
AGI21669/c
LOCUS      AGI21669      689 bp      DNA      linear      GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
ACCESSION  AGI21669
VERSION     AGI21669.1 GI:16650834
KEYWORDS   GSS: GSS (genome survey sequence).
SOURCE     Pan troglodytes male lymphoblast DNA, clone: lib.PTB Chimpanzee Male
           Pan troglodytes
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE  1 (sites)
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
           Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of library PTB
JOURNAL    2 (bases 1 to 689)
REFERENCE  Unpublished
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
           Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
REFERENCE  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
AUTHORS    1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
           Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
PRIMERS
LIBRARY    Sequencing: -21M13
VECTOR     : PKS145
R.Site 1   : SacI
R.Site 2   : SacI.
LOCATION/Qualifiers
1. 689
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-130M15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 164 a 159 c 176 g 189 t 1 others
ORIGIN
Query Match 33.0%; Score 439.2; DB 12; Length 689;
Best Local Similarity 87.0%; Pred. No. 4.2e-115;
Matches 507; Conservative 0; Mismatches 66; Indels 10; Gaps 2;

```

```

OY 556 caggaactgctggaattccctagagctgaactaagaatcccaagaagcttancgtggaagtgc 615
DB 366 CAGGACTATCTGGATTTCCTAGAGCCAACTAAGATTCCTTAAGCTTACTGGGAAGGTGAC 307
OY 616 cgcattccatttaaaacatggtggttgcaacttagctcacaccgcaatc----- 667
DB 306 CACATTCATTTTAAACACAGAGGCTTCGCACTTAGCTACACCCGAGCAATCAGTAGTA 247
OY 668 -agaagactactaaatgctactaagcaaaaacagaggaatgaagcaatgcaatcat 726
DB 246 AAGAGGGCTCACTAAATGCTTAATTAGCAAAAACAGAGCTAAAGAAATAGCCAAATAT 187
OY 727 ctattgctcctagagacagaggggaaggaacagagattggatataaactcaggtcacaag 786
DB 186 TTATTGCTTGAAGCTACAGCGGAGAGCAATGATCGGATATTAATCATGCAATTCAG 127
OY 787 ccagcaacagcaaccccccttgggtccctcccatgtatgg 829
DB 126 CCACCAATGGCTACCCCTTTGGGTCCTCCCTTGTGTATGG 84
RESULT 10
BHI49565/c
LOCUS      BHI49565      921 bp      DNA      linear      GSS 27-AUG-2001
DEFINITION ENTQ48tr Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION  BHI49565
VERSION     BHI49565.1 GI:15310303
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 921)
AUTHORS    Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL    HM1:IMSS sheared DNA library (2001)
COMMENT    Contact: Brendan J Loftus
           Department of Eukaryotic Genomics
           The Institute for Genomic Research
           9712 Medical Center Dr., Rockville, MD 20850, USA
           Tel: 301 838 0208
           Fax: 301 838 3543
           Email: b.loftus@tigr.org
           Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
           DNA library
           Seq primer: M13-Reverse
           Class: shotgun
           High quality sequence start: 40
           High quality sequence stop: 567.
FEATURES
source
1. 921
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + l method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 242 a 216 c 226 g 237 t
ORIGIN

```

Query Match 33.0%, Score 438.8; DB 12; Length 921;
 Best Local Similarity 83.8%; Pred. No. 6.3e-115;
 Matches 550; Conservative 0; Mismatches 84; Indels 22; Gaps 4;

547 actgagaaacaggaactgaactgattctcttaagctgaactaaagctcnaagctcncg 606
 |||||||
 651 AGTGAGAGACAGAGACTACTGATATCTTAAAGCCCTAAGATCTTAAAGCTAGCG 592
 |||||||
 607 gaaggtgacccatccatcttaacacaggggcttgcacactgaagctcaccacccacat 666
 |||||||
 591 GAAGGTGACTGATCCACTTAATAACAGGGGCTTGCAACTTACCTACACCCGACAT 532
 |||||||
 667 c-----agagagctcaactaaatgctcaatcagcaaaaacagaggttaagcaata 717
 |||||||
 531 CAGCTAGTAAAGAGAGCTACATAAAACCTAATATAGCAAAACAGAGAGTAAAGAAATA 472
 |||||||
 718 gccaatctatctgctcggagagacagcggaagagacagatgagatataaactcag 777
 |||||||
 471 GCCAATCATCTATCCCTGAGAGACACTGGAGGAGACAAATGA-TAGATATTAACCCAG 413
 |||||||
 778 gcaatcaagccagcaacagcaaaccccttggctccctccatgtataggagctcgt 837
 |||||||
 412 GCATTCAACACGACAGAGGCTAACCTTTTGAAGTCCCTCTTGTATGGAGCTCTGT 353
 |||||||
 838 ttcaactatctcaactatataaatacagactg--caactctctggtccggtgtttc 895
 |||||||
 352 -----TTTCACTATTAATCTTGCAACTGCACACACTCTTGCTGTCAGTGTGTGT 303
 |||||||
 886 tatgctcaagctggtgtttgttgcacatcacacactgctgtttgccacgctcaaac 955
 |||||||
 302 TACGCTGAGCTGAGCTTTTCACTTGGCTCCACACAGCTGCTTGGCCGCAATCCCAAC 243
 |||||||
 956 ccgcgtcgtgaactcaccctcttggatcagacagagtgctcagctgtgctcagctcagc 1015
 |||||||
 242 CCGCGCGTGAAGCTCCACCCCTCCAGATCTGGCAGGGTGTCTGCTGATCCGATCCAGT 183
 |||||||
 1016 gaggtacacattgcacatcccgatcagactaaagcttgcacatgttcttcgcatgctaa 1075
 |||||||
 182 GAGGACCCCATTTGGCTGTGATGGCTTAAGCTGTGTGCTGTGACAGGCTTAA 123
 |||||||
 1076 gtgcctggtgtgtccttaataagaactgaacactgtcactgtgtcctgttctcttc 1135
 |||||||
 122 GTGCCGGGTCTATCTATCTAATGAGCTGAACATTAAGTCTGGGTCCACAGTCTCTCT 63
 |||||||

Db 1136 atgacccacaggtcttcaataagactataaacatccacagctgagcccaagatccat 1191
 |||||||
 62 GAGACCCACGCGCTTAAATAGAGCTAATACGACGACTGCATGGCCCAAGATTCCAT 7
 |||||||

RESULT 11
 AG076758/c 679 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-071A23.R, genomic survey sequence.
 DEFINITION AG076758
 ACCESSION AG076758
 VERSION AG076758.1 GI:16628560
 KEYWORDS GSS; GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
 BAC library clone: PTB-071A23.R.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (sites)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 679)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suenhiro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chimpbes@gsr.riken.go.jp, URL: http://ngp.gsr.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.

COMMENT PRIMERS
 Sequencing: M13Rev
 LIBRARY
 Vector : pRS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..679
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-071A23.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"
 BASE COUNT 163 a 157 c 187 g 171 t
 ORIGIN

Query Match 32.8%, Score 435.6; DB 12; Length 679;
 Best Local Similarity 86.3%; Pred. No. 4.6e-114;
 Matches 480; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

133 aataatgtagagagagaggaactctcaaaacactcacccctggggccctcagcaaatg 192
 |||||||
 599 AACTCAATGTAGAGATGATGATTTCAAGACGCGCGCTATTTGGGCTGTAGCAATG 540
 |||||||
 193 gatgcctcgtacatccctctctcttagaacctcagaactaataatcttactcctct 252
 |||||||
 539 GATGCCCTGGGTGCGCCCTCTTGAACCTGTAGACACTGTGAAGGCTGCTCTCTT 480
 |||||||
 253 tggacctgtatctcaactctccttgtaagtgtgtctctccagaatgaagctgttaa 312
 |||||||
 479 TGGGCCCTGTATATTGACCTCTCTTGTGAGTTGTCTTCCAGAAATTAAGGCTTAA 420
 |||||||
 313 gctacaatagttcttcaatgaaacccagaatgacatgaactgaactaaatcactgctg 372
 |||||||
 419 GCTACAAATGTTCTTCAAAATGAAGCCCAATGACATGCTAATTAATTAACATGT 360
 |||||||
 373 acccctgacccgctcgtctagactatgctctgtaagttaatgaactgaatcaccctcc 432
 |||||||
 359 ACCCCTGGGCCAGCCTCTAGGCCCTTGGTCCGATGTATGACATCAAAAGGACCCCTCC 300
 |||||||
 433 cgaggaatctcaactctgacaaacccctactacaccccaatcagtagaagagtagag 492
 |||||||
 299 TGAGGAATCTCTAATCTGACAAACCCCTACATGCCCATTTTCAGCAGGAGAGGATAG 240
 |||||||
 493 caattgtcagccaaaccccccacagtaactgtggtgttccctgttgaaggggtgagtag 552
 |||||||
 239 CGGTGCTGATGACACTCTCCAGACGACTTGCGTTTCTGTTGAAGAGGGGTACTGAG 180
 |||||||
 553 agacagagactagctggaattctcctagcctgactaagaatcccaagcctantctggagag 612
 |||||||
 179 AGACAGGAGACTGATGATTTCTTAGACGAGCTAAGAAATCCCTTAACCTTAAGCTGGAAG 120
 |||||||
 613 gaccgcatccatcttcaactatgaggtcttgcaacttaactcaacccggaactaagaga 672
 |||||||
 119 GACCCGCTCAACCTTTAAACACGGGGCTTGGCAATTAAGCTACACCCGACCAATAGAGA 60
 |||||||
 673 gctcactaaatgcta 688
 |||||||
 59 GCTCTGAGCATGCTTA 44
 |||||||

RESULT 12
 AG102951 683 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-106616.F, genomic survey sequence.
 DEFINITION AG102951
 ACCESSION AG102951

TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other SSS: RPII11-64L19.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamast@tigr.org

Clones are derived from the human BAC library RPII-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufile.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufile.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="GDB:7524474"
/db_xref="taxon:9606"
/clone="RPII-11-64L19"
/clone_lib="RPII-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPII11 Human Male BAC Library"

BASE COUNT 183 a 211 c 154 g 166 t
ORIGIN

Query Match 32.2%; Score 427.6; DB 12; Length 714;
Best Local Similarity 83.0%; Freq. No. 9.4e-112;
Matches 565; Conservative 0; Mismatches 92; Indels 24; Gaps 6;

```

OY 589 atccnaagccttact-gggaagtgacccatccatcttaaacatggggcttgaact 647
DB 1 ATTCCCTAAGCCTTAGTGGGAGGTGACACACCCTCTTTAAACACAGAGCTTGTACT 60
OY 648 tagctcacaccgacacatc-----agagagctcactaaatgtaatacagc-aa 697
DB 61 CAGCTCACACCCGACCTATCGAGTAAGAGAGCTCACTAAATATACCAATTAGGCTAA 120
OY 698 aacacagaggttaagaacat-agccaatcatctatctgctgagagacagcggaagaca 756
DB 121 AAACAGAGAGGTAAAGAAATATCAATCATCTATCGCTGAGACACAGGGGAGGACA 180
OY 757 aggatgggataataaactcagagctcaagcaagcaaaccccttgggtccct 816
DB 181 ATGATCGGATATTAACCCAGGCACTTGAGCCAGATCAGTAACCTCTTTGGGTCCCT 240
OY 817 cccattgtagggagctctgttttcaactatctatctatctatctatcaatcgaactg--c 874
DB 241 CACACTGATATGGAGCTCTGT-----TTTCACTGATTAATCTTGCAACTGCAC 290
OY 875 acctctcgttcggtgtttttatgctcaagctgagcttltgtcgccatccacactg 934
DB 291 ACTCTTCTGGTCAATGTTTTCGGCTCAAGCTGAGCTTTGCTCCCGCCACCACTG 350
OY 935 ctgttgcacacgctcaagacccgctgctgaactcactcccttggatccagcagagtg 994
DB 351 CTGAATGCCCGCATGCGAGACTGCTCTGACTTCCACCCCTCGGATCCGGCAGAGTGT 410
OY 995 ccaactgtctcctgatacagcagagtaaccattgcaactcccgatcaggtctaaagctg 1054
DB 411 CCCCTGCATCTCTGATCAGAGGACCCCATTTGCCACTCCGATCAGGCTAAAGGCTTG 470
OY 1055 ccaattgtcctgatagtcgaagtgccgtgttgccttaatagaactgaacactggtcac 1114
DB 471 CCATTGTTCTCTGCACAGCTAAGTCTGGGTCTCACTTAATCAGGCTGGAACACTGTTCCG 530

```

```

OY 1115 tgggttccatggtctctctccatgacccacggtcttcaatagactataaacactaccgc 1174
DB 531 TGGGTTCCACAGGTTCTCTTCCATGACTCACAGCTTCTTAATAGACTTAACACTTACCAC 590
OY 1175 atggcccaagatccatctccttggatctgtgagcccaagaacccaggtcagagaagt 1234
DB 591 ATGCCCCAGGTTCCATTCCTTGGAATCCATGATGCCAAGAACCCCGTCAGAGAAATA 650
OY 1235 gaggtctgcacacattggga 1255
DB 651 AAGGCCCGCCCATCTTGGGA 671

```

Search completed: June 20, 2002, 08:18:47
Job time: 7373 sec

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

```
Title: US-09-319-156A-12
Perfect score: 1329
Sequence: 1 tcaaatcgaaagccttag.....gcaagatccccagtaaca 1329
```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

total number of hits satisfying chosen parameters: 767066

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :      Issued_Patents_NA.*
1 : /cgn2.6/ptodata/1/ina/5A.COMB.seq.*
2 : /cgn2.6/ptodata/1/ina/5B.COMB.seq.*
3 : /cgn2.6/ptodata/1/ina/6A.COMB.seq.*
4 : /cgn2.6/ptodata/1/ina/6B.COMB.seq.*
5 : /cgn2.6/ptodata/1/ina/PCrUIS.COMB.seq.*
6 : /cgn2.6/ptodata/1/ina/PCrUISseq.COMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	731.8	55.1	2946	4	US-09-175-928-3	Sequence 3, Appl1
2	208	15.7	279	4	US-08-686-878-3	Sequence 50, Appl1
3	208	15.7	279	1	US-08-721-489-4	Sequence 4, Appl1
4	166	12.5	1722	1	US-08-691-663-58	Sequence 58, Appl1
5	165	12.4	1859	3	US-08-691-563-46	Sequence 46, Appl1
6	161.8	12.2	80246	4	US-09-078-994-4	Sequence 4, Appl1
7	151.8	11.4	80595	4	US-09-078-994-3	Sequence 3, Appl1
8	80.6	6.1	1704	1	US-08-007-828-1	Sequence 1, Appl1
9	80.6	6.1	2518	4	US-09-011-745-1	Sequence 1, Appl1
10	80.6	6.1	5865	4	US-09-011-745-8	Sequence 8, Appl1
11	57	4.3	7218	1	US-08-232-653-14	Sequence 14, Appl1
12	50.8	3.8	8060	4	US-08-766-528-1	Sequence 1, Appl1
13	49.2	3.7	8132	4	US-08-766-528-3	Sequence 3, Appl1
14	46.4	3.5	1911	1	US-08-358-420-8	Sequence 8, Appl1
15	46.4	3.5	1965	1	US-08-358-420-9	Sequence 9, Appl1
16	46.4	3.5	3925	4	US-09-011-745-9	Sequence 9, Appl1
17	46.4	3.5	8202	1	US-08-358-420-13	Sequence 13, Appl1
18	44.8	3.4	6028	4	US-09-011-745-5	Sequence 5, Appl1
19	44.6	3.4	1989	4	US-09-332-278A-5	Sequence 5, Appl1
20	44.6	3.4	2499	1	US-08-105-483-30	Sequence 30, Appl1
21	44.6	3.4	2459	1	US-08-709-209-310	Sequence 310, Appl1
22	44.6	3.4	2459	1	US-08-458-101-310	Sequence 310, Appl1
23	44.4	3.3	10970	3	US-08-716-351A-5	Sequence 5, Appl1
24	44.2	3.3	8655	3	US-09-075-272-1	Sequence 1, Appl1
25	43.2	3.3	2462	3	US-09-111-085-1	Sequence 1, Appl1
26	43.2	3.3	7333	4	US-08-766-528-2	Sequence 2, Appl1
27	41.6	3.1	2001	3	US-08-850-961-13	Sequence 13, Appl1

45	37	2.8	4273	3	US-08-795-430-3	Sequence 13, Appl
44	38	2.9	8525	3	US-08-795-430-3	Sequence 1, Appl
43	39	2.9	8373	5	PCW-US93-08041-6	Sequence 2, Appl
42	39	2.9	8373	4	US-09-433-332B-2	Sequence 1, Appl
41	39	2.9	8373	2	US-08-886-642-8	Sequence 7, Appl
40	39	2.9	8373	1	US-08-110-300A-8	Sequence 6, Appl
39	39	2.9	1914	1	US-08-258-420-7	Sequence 30, Appl
38	40	3.0	6061	4	US-09-011-745-6	Sequence 25, Appl
37	40	3.0	3482	3	US-09-111-085-3	Sequence 24, Appl
36	40	3.0	2000	4	US-09-376-781-20	Sequence 23, Appl
35	40	3.0	2000	4	US-09-376-781-25	Sequence 22, Appl
34	40	3.0	1453	4	US-09-376-781-24	Sequence 21, Appl
33	41.2	3.1	6332	4	US-09-011-745-7	Sequence 20, Appl
32	41.6	3.1	6332	4	US-09-479-776-1	Sequence 19, Appl
31	41.6	3.1	5376	3	US-08-850-963-1	Sequence 18, Appl
30	41.6	3.1	5376	3	US-08-073-836-2	Sequence 17, Appl
29	41.6	3.1	5109	1	US-08-073-836-1	Sequence 16, Appl
28	41.6	3.1	2001	4	US-09-479-776-13	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-175-928-3
; Sequence 3, Application US/09175928A
Date: 03/03/2009

GENERAL INFORMATION:

APPLICANT: McCoy, John M.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl
APPLICANT: Merbern David

APPLICANT: Treacy, Maurice

APPLICANT: Genetics Institute, Inc.

FILE REFERENCE: 6006B.AJ172A

CURRENT FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 62

; SEQ ID NO 3

TYPE: DNA

ORGANISM: H
IS-09-175-928-3

Query Match 55.18; Score 731.8; DB 4; Length 2946;
Post Local Similarity 60.78; Pos 375-390

Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1.

QY 1 tcaaatcgaagagctttagacttgcctaaccgccaacaaadgaaggaaggaacctgtttatttt 60

Db 2067 tcaaaatcgaagaccttataacttaactaacccgtaaaaagagggaaacctatttat+++ 2126

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

121 aaLcGagAACCAACGcagagacCCcCaAaAcACcGcACCCcGgGcCT 180 :

218/ aattcgagatcgaatcacaaagttagagcagagagcttcgaaacactggaacctgggcct 2246

QY 181 cctcagccaatgtagtgccttgactctcccttcttaggaccttagcagctataatt 240

Db 2247 cctcagccaatgtagccctgatatcccccctcttagaacctctagcaqctataatat 2306

Ov 241 ttactcctcttgaacctatatctcaacttcctatataatctctccgaat 300

[illegible]

```

QY 301 tgaagcttgaagctacaagaatagttctcaaaatggaacccagatgacagctcagactaa 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2367 cgaagctgtaaacta-----caaatgagcccaagatgacagctcagactaa 2414
QY 361 aatctacccgtgagcccccgtgacccgctcgtctgaagactaagctctatgttaatgacatga 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2415 gatctaacgcagaccccccgtgacccgctcgtctgaagactaagctctatgttaatgacatga 2474
QY 421 agtcaaccccccgtgacccgctcgtctgaagactaagctctatgttaatgacatga 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2475 agtcaaccccccgtgacccgctcgtctgaagactaagctctatgttaatgacatga 2534
QY 481 aagcagttgagcagctgtgtcagccaaacctccccaagctactgtgtgttttccgttgaaga 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2535 aagcagttgagcagctgtgtcagccaaacctccccaagctactgtgtgttttccgttgaaga 2594
QY 541 ggtgtgagctgaggaacaggaatagctgatttctctagctgtcagtaagaatcccaagcct 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2595 tgggtgagctgaggaacaggaatagctgatttctctagctgtcagtaagaatcccaagcct 2654
QY 601 aactgggaaggtgagccagctcattcttaaacatggggtgtgcaactagctcagccg 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2655 agctgggaaggtgagccagctcattcttaaacatggggtgtgcaactagctcagccg 2714
QY 661 accaatcagaagagctcactaaatgtctaatcaaggcaaaacaggaaggtaaagcaatagcc 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2715 accaatcagaagagctcactaaatgtctaatcaaggcaaaacaggaaggtaaagcaatagcc 2774
QY 721 aatcatctatgctctgagagacagacagcggaaggaagatgtgagatataaaccagca 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2775 aatcatctatgctctgagagacagacagcggaaggaagatgtgagatataaaccagca 2834
QY 781 tcaagcagcaagaacagcaaccccttgggtccctcccatgtgataggagctctgtttt 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2835 ttcgagccggaagaagcaaccccttgggtccctcccatgtgataggagctctgtttt 2894
QY 841 caactatcttcaactctatataatcatgcaactgca 875
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2895 catgcatcttcaactctatataatctgcaactgca 2929

```

```

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: Linear
MOLECULE TYPE: cDNA
US-08-686-878A-50

Query Match      15.7%: Score 208; DB 1; Length 279;
Best Local Similarity 85.1%: Pred. No. 2,46-61;
Matches 223; Conservative 1; Mismatches 38; Indels 0; Gaps 0;

QY 614 accgcatccatcttaaacatggggtgtgcaactagctcagcccgacacagagag 673
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 RCGACATCCACTTTAAACACGGGNTTGCAANNAAGATNCACTTGACCAATCAGAGAG 60
QY 674 ctcaactaaatgctcaatcaaggcaaaacaggaaggtaaagcaatgccaatcatctatg 733
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 NTCANTAAATGATNATNTNGSCAAACAGAGGTAAGAAATAGCCAAATCATCTATTGC 120
QY 734 ctgagagacagcggaaggaagatgtgataataaccagcatcagccagcaaga 793
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CTGAGAGCAGCAGAGGAGGAGATGATGCGGATTAATACCAAGTTTNGACCGGCA 180
QY 794 cagcaaccccttgggtccctcccatgtgataggagctctgttttcaactatctac 853
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CGGCAACCCCTTTGGGTGCGCCCTTCCTTTGATGAGACTGTGTTTATGCTATTTCAN 240
QY 854 tctatataatcatgcaactgca 875
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TNFATTAATNTGCAACTGCA 262

RESULT 3
US-08-721-489-4
Sequence 4, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racine, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530

```



```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-721-489-4

```

```

Query Match      15.7%; Score 208; DB 1; Length 279;
Best Local Similarity 85.1%; Pred. No. 2,4e-61;
Matches 223; Conservative 1; Mismatches 38; Indels 0; Gaps 0;

```

```

Db      614 accgcatcatttaacagtgaggcttgcaactctcacaccgaccatcagag 673
      1 KCCACATCCCTTTAAACAGGGGNTTGCANANAGATNACCTTGACCAATCAGAG 60
Qy      674 ctactaaatgtctaatcagcaaaacagaggtlaagaatagccatcatctatgc 733
      61 NTCANTAAATGATNATTNGCAAAGAGAGGTAAAGAAATGCCAATCATCTATTGC 120
Qy      734 ctgagagcacagcggaagaaagatttgatataaactcaggcatcgaagcagaa 793
      121 CTGAGAGCACAGCAGGAGGACATGATCGGATATTAACCAAGTTTNGAGCCGCCAA 180
Db      794 caagaaccccttgggtccccccttgataggagcctctgttccactctatc 853
      181 CGGCAACCCCTTTGGTCCCTCCCTTTGATGGAGCTNCTTTCATGCTATTTCAN 240
Qy      854 tctatataatcagcaactgca 875
      241 TMTATTAAATNTTGCACTGCA 262
Db

```

```

RESULT 4
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:

```

```

;
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: APB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-691-563C-46

```

```

Query Match      12.5%; Score 166; DB 3; Length 1722;
Best Local Similarity 60.5%; Pred. No. 2e-46;
Matches 333; Conservative 0; Mismatches 200; Indels 17; Gaps 3;

```

```

Qy      2 caaatgagagagcttagactctgctaacgccaagagggaggaactgtttattt 61
      547 CAATAACCGCCGAGGCCACACCTTCCTCAGGAAGAGGAGCTTCGACCTTCTTA 606
Qy      62 ggggaagaatgctgtatgatatcatcgaatcgaatcattactggaagttaaaga 121
      607 GGGGAAGAGTGTCTTTTACACTTACACAGTCAGGATAGTACGAGATGCCACTGGCAT 666
Db      122 attgagatcgaatataatgtagagcagagagacctcaaaacactgcaacctg 181
      667 TTACAGGAAGAGGCTTGATATACGACAAATGCTTCAAACTCTTATACCA-----C 720
Qy      182 ctgagcaatgagatgacctgagacctccctctcttagagacctcagaagctatatt 241
      721 CTCTGAGATGGGCAACATGAGCTTCTTCCATTCTAGCTCCATGCGACCATCTTGCTG 780
Db      242 ttaactcctcttgagacctgatatctcaactctctgttaagttgtctctcagaatt 301
      781 TTACTCACTTGGGCGCTGTATTTTAAGCTTCTTGCAAAATTTGTTCCCTAGAGATC 840
Qy      302 gaagctgtaaagctacaaatagttcttcaaatggaacccagatgagtcataactaa 361
      841 GAAGCCATCAAGCTTACAGATGCTTCAAAATGGAACCCCAAAAGATTCAACTAACAAAC 900
Db      362 atctacgttgagacctcgagccgctgtagactatgctgtgttaagacatgaa 421
      901 TTCTACCAAGGACCTGTGAACGATCCACTGCACTTCCACT-----ACCTAGA 950
Qy      422 gtcaacctcccgaggaatctcaactgacacacccctactacactccaattcagtaga 481
      951 GATTCCTCTGTGAAGACACTACACACTGCAAGGCGCTTTTGCCCTATCCAGCAGGA 1010
Qy      482 agcagttagagcagctgctcagccaaacctcccaacagtaactgttctctgttagag 541
      1011 AGTACTAGAGCGGCTCATCGGCAAA-ATTCCACAGAGAGTTGGGCTGCTGTTAGAG 1069
Db
Qy      542 ggtgactga 551
      1070 GGGGAGATTGA 1079
Db

```

```

RESULT 5
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND

```


Qy 242 ttaacttcctcttgagccctgtatctcaacttcctctgttaagtgtgcctccgaatt 301
 Db 3028 atactaacccattgagccattgcttttcagtcgcctcattgacctcaatatgatgaatt 3087
 Qy 302 gaagctgttaa 312
 Db 3088 atgtgtgtaca 3098

RESULT 11
US-08-232-463-14

APPLICANT: DORNER, F.
 APPLICANT: SCHEIFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS.
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:

```

ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109

```

? INFORMATION FOR SEID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 7218 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? CLONE: pTZgpt-F15
?
US-08-232463-14

```

Query Match	4.38;	Score 57;	DB 1;	Length 7218;
Best Local Similarity	5.88;	Pred. No. 1.2e-08;		
Matches 24;	Conservative 223;	Mismatches 168;	Indels 0;	Gaps 0;

Qy	729	atgcgcgaagacacacgcgggaagagacaagatgtagatataaaccagcattcaagc	788
Db	1026	ATTAAATTCGAGCGTTGGCTCGAGTCGAGGAGACTTCGCACTYYYYYYYYYYYYYYY	1085
Qy	789	agcaacagcaacccccctttgggttcctcccatctgataggagactctgtttcaactctat	848
Db	1086	YYY	1145
Qy	849	ttcaactcattaaatcatcgaactgcactctctcgtccgtgtttttatggtccaagct	908

Db 1146 yyy 1205
Qy 909 gacgttctgttcgcacacacactgtctttgcacgctcacagaccgcgtgactt 968
Db 1206 yyy 1265
Qy 969 ccaccccttctgatacagcagagtgctcaactgtctctctgataccagcgagtaeccatlg 1028
Db 1266 yyy 1325
Qy 1029 ccactccgcatacagctaaagcttgcattctccctgatactgaagtcctggtttg 1088
Db 1326 yyy 1385
Qy 1089 tctcaatagacttcaaacactgctcactggttcacatgctctctctcacaagccca 1143
Db 1386 yyytgcacCA 1440

RESULT 12
US-08-766-528-1

APPLICANT: Jay A. Fishman
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 74

ADDRESSSEE: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-POS
SOFTWARE: PatentIn Release #1.0, Version #1.235
CURRENT APPLICATION DATA:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/572,645
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:

```

1      REGISTRATION NUMBER: 35,965
2      REFERENCE/DOCKET NUMBER: M6P-038C
3      TELECOMMUNICATION INFORMATION:
4      TELEPHONE: (617)227-7400
5      TELEFAX: (617)227-5541
6      INFORMATION FOR SEO ID NO: 1:
7      SEQUENCE CHARACTERISTICS:
8      LENGTH: 8060 base pairs
9      TYPE: nucleic acid

```

MOLECULE TYPE: CDNA
US-08-766-528-1

Query Match	3.8%	Score	50.8	DB	4	length	8060
Best Local Similarity	55.7%	Pred. No.	1.7e-06				
Matches	97	Conservative	0	Mismatches	77	Indels	0
				Gaps	0		
QY	2	caaatcgaaagacgttgaactgctaaccgccaagaagagggggaacctgttatttta	61				
Db	1619	CAGATTAGACGAGGGCTTGACATTATATTATTTCTAAAGGAGGAGCTATAGTACGCTTG	1678				
QY	62	ggggagaagaatgctgttagtatgttaacaattctggaattactacttgaagaagttaaga	121				

Db 1679 AAGGAGGATGCTGTTTATGTCATTCATGAGGGCCATCAGACCTCCATGAACAA 1738
Oy 122 attgaatcgaataatgtagagcagagaccctcaaacacgtccaccctgg 175
Db 1739 CTTAGAGAAAGGTTGAGAGAGCGTCGAAGGGAAGAAAGAACTACTCAAGGCTGG 1792

RESULT 13

US-08-766-528-3

; Sequence 3, Application US/08766528

; Patent No. 6190861

; GENERAL INFORMATION:

; APPLICANT: Jay A. Fishman

; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD, LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/766,528

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/572,645

; FILING DATE: 14-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Louis Myers

; REGISTRATION NUMBER: 35,965

; REFERENCE/DOCKET NUMBER: MGP-038CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8132 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; -08-766-528-3

Query Match 3.7%; Score 49.2; DB 4; Length 8132;
Best Local Similarity 55.2%; Pred. No. 6e-06; Mismatches 78; Indels 0; Gaps 0;

Oy 2 caaatcgaagagcttgaactgtcgaacccgcaaaagagggggaacctgttatttta 61
Db 7153 CAGATAGAGAGGCGGTGATTATTATTCATAAAGAGAGGATATGTGTACCTTG 7212
Oy 62 ggggaagaatgctgttagatgtaataatcaatctggaatcattacatgagaagaatgaa 121
Db 7213 AAGGAGGAATGCTGTTTATGTCATTCAGGGGCCATCAGAGACTCCATGAACAG 7272
Oy 122 attgaatcgaataatgtagagcagagaccctcaaacacgtccaccctgg 175
Db 7273 CTTAGAGAAAGGTTGAGAGAGCGTCGAAGGGAAGAAAGAACTACTCAAGGCTGG 7326

RESULT 14
US-08-258-420-8
; Sequence 8, Application US/08258420
; Patent No. 5710037

; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; APPLICANT: Vanlo, Eljo F.
; TITLE OF INVENTION: No. 5710037e1 Retroviral Envelope and LTR and Retroviral Ve
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carrella, Byrne, Bain, Giffillan, Cecchi, Stewart & Olstein

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,420

; FILING DATE: 10-JUN-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Olstein, Elliot M.

; REGISTRATION NUMBER: 24,025

; REFERENCE/DOCKET NUMBER: 271010-208

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1911 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: retroviral envelope sequence

; US-08-258-420-8

Query Match 3.5%; Score 46.4; DB 1; Length 1911;
Best Local Similarity 48.1%; Pred. No. 2.1e-05;
Matches 177; Conservative 0; Mismatches 176; Indels 15; Gaps 1;

Oy 2 caaatcgaagagcttgaactgtcgaacccgcaaaagagggggaacctgttatttta 61
Db 1528 CAGAACCGCAGAGCGCTGATTGCTATTCCTAAGGAGGAGGCTCTGCGACCCCTA 1587
Oy 62 ggggaagaatgctgttagatgtaataatcaatctggaatcattacatgagaagaatgaa 121
Db 1588 AAGAGAGATGCTGTTTATGTCAGACACACAGCGGCTAGTGAGAGAGACATGGCCAA 1647
Oy 122 attgaatcgaataatgtagagcagagaccctcaaacacgtccaccctgggc--- 178
Db 1648 TTAAGAGAAAGGCTTAACAGACAAAAAAGTATTGAGACAGGCCAAGAGATGTCGAA 1707
Oy 179 -----ctccgaagcaatgtagtgcocctggaactccocctcttggaacctta 226
Db 1708 GGGCTGTTAATAGATCCCGCTGGTTTACACCTTAATCTCACCATATGAGACCTCTA 1767
Oy 227 gcaatcgaataatcttactcctcttggaacctgtaactcaactcctgttaagtt 286
Db 1768 ATAGTACTTACTGATCTTACTCTTTGAGCTTCCATTCATGATGATGATTT 1827
Oy 287 gctctccagaatgtagagctgtaagactacaatagttcttcaaatggaacccagaatg 346
Db 1828 GTTAAAGACAGGATATCAGTGTCCAGGCTAGTTTGACTCAACATATACACAGCTG 1887
Oy 347 cagttcat 354

Db 1888 AACCTAT 1895

RESULT 15

US-08-258-420-9

Sequence 9, Application US/08258420

Patent No. 5710037

GENERAL INFORMATION:

APPLICANT: Nienhuis, Arthur W.

APPLICANT: Vanin, Ello F.

TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vectors

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Carrella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein

STREET: 6 Becker Farm Road

CITY: Roseland

STATE: New Jersey

COUNTRY: USA

ZIP: 07068

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/258,420

FILING DATE: 10-JUN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Olstein, Elliot M.

REGISTRATION NUMBER: 24,025

REFERENCE/DOCKET NUMBER: 271010-208

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1965 bases

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

US-08-258-420-9

Query Match 3.5%; Score 46.4; DB 1; Length 1965;

Best Local Similarity 48.1%; Pred. No. 2.1e-05;

Matches 177; Conservative 0; Mismatches 176; Indels 15; Gaps 1;

QY 2 caaatcgaaagagcttagaactcttaacgcgcaaaagagggaacctgtttatttta 61
Db 1582 CAGAACCGCAGAGGCTGATTTGCTATTCCTAAGAGGAGGCTCTCGCAGCCTTA 1641
QY 62 ggggaagaatgctgttagtcaatcaatcgtgaatcattactgagaagtttaagaa 121
Db 1642 AAACAAGATGTTTATTATGACACACACGCGGCTAGTAGAGACAGCATGGCCAAA 1701
QY 122 attgagatcgaaatataatgtagcagaggaacctccaacaactgcaacctgggac-- 178
Db 1702 TTAAAGAAAGGCTTAATCAAGACAAAACCTATTGAGACAGGCCAAGATGGTTCGAA 1761
QY 179 -----ctcccgagcaatgtagtgcctggacctccctctcttagaactcta 226
Db 1762 GGGCTGTTAATAGATCCCTGCTTTACACCTTAATCTCACCACATCATGGACCTCTA 1821
QY 227 gcaagctataatatttactcctcttggaccctglatctcaactcctgtttaagttt 286
Db 1822 ATACTACTCTACTGATCTTACTCTTTGGACCTTGCAATTCATTCATTCATTCATTT 1881
QY 287 gtccctccagaattgaagctgtaagctacaataagtlcttcaatggaaccccaagatg 346

Db 1882 GTTAAAGACAGGATATCAGTGTCCAGGCTCTAGTTTGACTCAACAAATATCACACGCTG 1941

QY 347 cagtcacat 354

Db 1942 AACCTAT 1949

Search completed: June 20, 2002, 09:40:25
Job time: 8535 sec

THIS PAGE BLANK (USPTO)

PR 26-NOV-1996; 96US-0756429.
 XX (INMR) BIO MERIEUX.
 XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
 PI Mandrand B, Paranhos-Baccala G, Perron H;
 XX MPI: 1998-322732/28.
 DR P-PSDB: AAM71069.
 XX
 PT New nucleic acid from retroviruses - useful for diagnosis,
 prevention and treatment of, e.g. multiple sclerosis
 PS
 XX Disclosure: Pages 187-188; 286pp; English.
 CC The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) genomic fragment used in the method of the
 CC invention. The invention provides complete or partial genomic
 CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
 CC polypeptides encoded by these genes. The invention also provides
 CC antibodies raised against the polypeptides. The genomic sequences,
 CC polypeptides and antibodies are also claimed useful for diagnosing
 CC infection by MS and rheumatoid arthritis-associated viruses, and also
 CC for prevention and treatment of infection with these viruses.
 XX
 SO Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 99.8%; Score 1326; DB 19; Length 1329;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcaaaatcgaagaagcttagacttgcgaacgcaaaagaggaggaacgtttatttt 60
 Db 1 tcaaaatcgaagaagcttagacttgcgaacgcaaaagaggaggaacgtttatttt 60
 QY 61 agggagaagaatgctgtatgtatgaatcaatcgtgaatcattactgagaagttaaaga 120
 Db 61 agggagaagaatgctgtatgtatgaatcaatcgtgaatcattactgagaagttaaaga 120
 QY 121 aattggaatcgaatatatgtagagcagagaggaacgttcaaaacactgtgacgtggagct 180
 Db 121 aattggaatcgaatatatgtagagcagagaggaacgttcaaaacactgtgacgtggagct 180
 QY 181 cctcagcaatgtagtgccttgagacttccctctttagaactcttagagctatgaatatt 240
 Db 181 cctcagcaatgtagtgccttgagacttccctctttagaactcttagagctatgaatatt 240
 QY 241 tttaactctcttgagacccctatctcaacttctcttgtaagtgtgtctctccaagat 300
 Db 241 tttaactctcttgagacccctatctcaacttctcttgtaagtgtgtctctccaagat 300
 QY 301 tgaagctgtaaagctacaataatagctctcaaatgtaaacccagatgacgttcaatc 360
 Db 301 tgaagctgtaaagctacaataatagctctcaaatgtaaacccagatgacgttcaatc 360
 QY 361 aatctacgttgagacccctgagcgcctgctagactatgactgtatgtaatgacattga 420
 Db 361 aatctacgttgagacccctgagcgcctgctagactatgactgtatgtaatgacattga 420
 QY 421 agtccacccctccgaggaatctcaactgcaaacccctactaacctcaatttcgtatgg 480
 Db 421 agtccacccctccgaggaatctcaactgcaaacccctactaacctcaatttcgtatgg 480
 QY 481 aagcagcttagagcaagtctcagcaacactcccaacagtaactgtgttcttcctgttga 540
 Db 481 aagcagcttagagcaagtctcagcaacactcccaacagtaactgtgttcttcctgttga 540
 QY 541 ggggtgacttgagagacagagctagcttgatctcctaaagctgactaagaatccnaagcct 600
 Db 541 ggggtgacttgagagacagagctagcttgatctcctaaagctgactaagaatccnaagcct 600
 QY 601 aactggaagtgtagcgcacatcattttaaacaatgggctgtgcaactagctcaacccg 660

Db 601 aactggaagtgtagcgcacatcattttaaacaatgggctgtgcaactagctcaacccg 660
 QY 661 accaactcagagagctcactaaatgctactaagcgaacaaacagagatgaagcaatagcc 720
 Db 661 accaactcagagagctcactaaatgctactaagcgaacaaacagagatgaagcaatagcc 720
 QY 721 aatcatctatgtccttgagagacagcagcggaagagacaagatctggatataaactcaagca 780
 Db 721 aatcatctatgtccttgagagacagcagcggaagagacaagatctggatataaactcaagca 780
 QY 781 ttcaagccaagacaagcaaccccttgggtccctcccatctgatactggagctcgtttt 840
 Db 781 ttcaagccaagacaagcaaccccttgggtccctcccatctgatactggagctcgtttt 840
 QY 841 cactctatctactatataatcatgtaactgtgcaactctctctgtcctgtttttatgg 900
 Db 841 cactctatctactatataatcatgtaactgtgcaactctctctgtcctgtttttatgg 900
 QY 901 ctcaagctgagctttgttgcgcacatccacactgctgtttgcacacgttcaacagaccgct 960
 Db 901 ctcaagctgagctttgttgcgcacatccacactgctgtttgcacacgttcaacagaccgct 960
 QY 961 gctgacttccatcccttggatccagcagagtgctcactgtgtctctatccagagagt 1020
 Db 961 gctgacttccatcccttggatccagcagagtgctcactgtgtctctatccagagagt 1020
 QY 1021 acccaactcgaactccgcatcagagctgcaatgtctctcgtcactgagctaaagtgcc 1080
 Db 1021 acccaactcgaactccgcatcagagctgcaatgtctctcgtcactgagctaaagtgcc 1080
 QY 1081 tgggttctccttaataagaactggaacactgtgcaactgtgttccatgtgtctctccatgac 1140
 Db 1081 tgggttctccttaataagaactggaacactgtgcaactgtgttccatgtgtctctccatgac 1140
 QY 1141 ccaagcgtcttaataagaacttaacactcaccggaatggcccaagaattcattcctctgta 1200
 Db 1141 ccaagcgtcttaataagaacttaacactcaccggaatggcccaagaattcattcctctgta 1200
 QY 1201 tctgtgaagcgaagaacccaggtcagagaangtgaagcttgcacacatttggggaatgg 1260
 Db 1201 tctgtgaagcgaagaacccaggtcagagaangtgaagcttgcacacatttggggaatgg 1260
 QY 1261 cccactgtcatttggtagcgccacacacatcttggagactgtggagcaagatccc 1320
 Db 1261 cccactgtcatttggtagcgccacacacatcttggagactgtggagcaagatccc 1320
 QY 1321 ccagtaaca 1329
 Db 1321 ccagtaaca 1329

RESULT 2
 AAX29704
 ID AAX29704 standard; DNA; 1329 BP.
 XX
 AC AAX29704;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE Clone 5M6 from MSRV-1.
 XX
 KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KM rheumatoid polyarthritis; ss.
 OS
 XX Multiple sclerosis related virus type 1.
 PN FR2765588-A1.
 XX
 PD 08-JAN-1999.
 XX
 PF 07-JUL-1997; 97FR-0008816.
 XX

```
PR 07-JUL-1997; 97F8-0008816.
XX
XX (INMR) BIO MERIEUX.
XX
XX WPI: 1999-098275/09.
XX P-PSDB: AAW9554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritis
XX
XX Claim 1: Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with
XX rheumatoid polyarthritis.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;
SQ
Query Match 99.8%; Score 1326; DB 20; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcaaaatcgaaagccttagactgtctaacgcgcaaaagaggggaacctgtatttt 60
DB 1 tcaaaatcgaaagccttagactgtctaacgcgcaaaagaggggaacctgtatttt 60
QY 61 aggggaagaatgctgtatgtatgttaatcaatcttggaatcattacttgaagaatlaaga 120
DB 61 aggggaagaatgctgtatgtatgttaatcaatcttggaatcattacttgaagaatlaaga 120
QY 121 aatttagatcgataataatgtagagagagaccccttcaaaacacgtgcacctggggcct 180
DB 121 aatttagatcgataataatgtagagagagaccccttcaaaacacgtgcacctggggcct 180
QY 181 cctcagcaaatgatgacctggagactcctccctcttagagaccttagacgataatatt 240
DB 181 cctcagcaaatgatgacctggagactcctccctcttagagaccttagacgataatatt 240
QY 241 ttactcctcttggagacctgtatctcaactcctctgttaagtgtgtctcttcagaaat 300
DB 241 ttactcctcttggagacctgtatctcaactcctctgttaagtgtgtctcttcagaaat 300
QY 301 tgaagctgttaagctataaataatttctcaaatggaaaccccaatgtagtccatgactaa 360
DB 301 tgaagctgttaagctataaataatttctcaaatggaaaccccaatgtagtccatgactaa 360
QY 361 aatctaacgtggagaccttgagccgctgtgtagactatgtctcgatgttaatgaattga 420
DB 361 aatctaacgtggagaccttgagccgctgtgtagactatgtctcgatgttaatgaattga 420
QY 421 agtcaacctcccgaggaatctcaactgycacacccctactacactccaattcaatgtag 480
DB 421 agtcaacctcccgaggaatctcaactgycacacccctactacactccaattcaatgtag 480
QY 481 aagcagtttaagcagatgttagcacaactccccaacagtagtcttggtttcctgttaga 540
DB 481 aagcagtttaagcagatgttagcacaactccccaacagtagtcttggtttcctgttaga 540
QY 541 gggctgagctagagagacgagctagctgattcttccttagagcgcgactaaatcccaagcct 600
DB 541 gggctgagctagagagacgagctagctgattcttccttagagcgcgactaaatcccaagcct 600
QY 601 aactgaggaaggtgacgcgacatctttaaacaatgggctgtgcaacttagctcacaccg 660
DB 601 aactgaggaaggtgacgcgacatctttaaacaatgggctgtgcaacttagctcacaccg 660
QY 661 accaatcagagagctcctcaataatgctcaatcagcgaacaaacaggaagttaaagcatagcc 720
DB 661 accaatcagagagctcctcaataatgctcaatcagcgaacaaacaggaagttaaagcatagcc 720
```

```
QY 721 aatcattatgtcctgtagagacagcgaggaagacaagaattggatataaactcagga 780
DB 721 aatcattatgtcctgtagagacagcgaggaagacaagaattggatataaactcagga 780
QY 781 ttaagccagacaagaacaccccttgggtccctccatctgtatgagagctctgttt 840
DB 781 ttaagccagacaagaacaccccttgggtccctccatctgtatgagagctctgttt 840
QY 841 cactctattcactctatataatcaatgacatgacatctctctgtcgtgttttttag 900
DB 841 cactctattcactctatataatcaatgacatgacatctctctgtcgtgttttttag 900
QY 901 ctcaagctgagcttctgttcgcacatccacacacgtgttggcaccgctcacagaccgct 960
DB 901 ctcaagctgagcttctgttcgcacatccacacacgtgttggcaccgctcacagaccgct 960
QY 961 gctgacttccatcccttggatccagagagtgltccactgtgctcctgtatccagcgaagt 1020
DB 961 gctgacttccatcccttggatccagagagtgltccactgtgctcctgtatccagcgaagt 1020
QY 1021 acccattggcactcccgatcaggtcgaagcttgccatttctctgtagtgaatggcc 1080
DB 1021 acccattggcactcccgatcaggtcgaagcttgccatttctctgtagtgaatggcc 1080
QY 1081 tgggttgccttaataagaactggaacactgtgcaactgggttccatggtctctccatgac 1140
DB 1081 tgggttgccttaataagaactggaacactgtgcaactgggttccatggtctctccatgac 1140
QY 1141 ccaaggtcttaataagaactggaacactgcaactgggttccatggtctctccatgac 1200
DB 1141 ccaaggtcttaataagaactggaacactgcaactgggttccatggtctctccatgac 1200
QY 1201 tctgtgagggcgaagaaccccgatcagagangtgaagcttgcacacatttgggaatgg 1260
DB 1201 tctgtgagggcgaagaaccccgatcagagangtgaagcttgcacacatttgggaatgg 1260
QY 1261 cccactgccatttctgtgtagcgccacacacatcttggagctgtgtaggaagatccc 1320
DB 1261 cccactgccatttctgtgtagcgccacacacatcttggagctgtgtaggaagatccc 1320
QY 1321 ccagtaaca 1329
DB 1321 ccagtaaca 1329
RESULT 3
AAS68626
ID AAS68626 standard; cDNA; 7466 BP.
XX
XX AAS68626;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #4430.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-0508631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
XX
```


PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
PA (INCYTE) INCYTE GENOMICS INC.

Chen SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
D'Sa SA, Ansley S, Dahl CR, Dam TC, Daniels SE;
Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiner TK, Daffo A;
Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX MPI: 2001-502867/55.
XX P-PSDB: AA019431.

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics

XX Claim 1: page 304; 522pp; English.

XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AA530986-AA531196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.

XX Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 58.6%; Score 778.8; DB 22; Length 1393;
Best Local Similarity 87.2%; Pred. No. 6,7e-245;
Matches 904; Conservative 0; Mismatches 112; Indels 21; Gaps 4;

QY 302 gaagctgaagctacaatagttcttcaatggaacccagagtgagctcactgaactaa 361
DB 1 gaagctgaagctacaatagttcttcaatggaacccagagtgagctcactgaactaa 60

QY 362 atctacgctgagccaccctgagccgctgctagactaactgctctgattgtaatgaactaa 421
DB 61 attacacagagaccctcgagccgctgctgctgagccactgacagctgattgtaattgatacga 120
QY 422 gtcaaccctccgaggaatctcaactgcaaacccctactacactlccaaattagtaga 481
DB 121 agcactcc-cccaaggaattcaactgcaaacccctactacactlccaaattagtaga 179
QY 482 agcagttagagcaggttgtagcgcgaacctcccaacagtaattggtttctctttagag 541
DB 180 agcagttagagcaggttgtagcgcgaacctcccaacagtaattggtttctctttagag 229
QY 542 ggtcagctgagagacaggaactagctgattctcttagctgtaactaaagatcccaagcta 601
DB 240 cgggagctgagagacaggaactagctgattctcttagctgtaactaaagatcccaagcta 239
QY 602 nctgggaagtgagccgcatccattaaacatggggcttgcaactagtcacaccca 661
DB 300 gctgggaagtgagccgctgctactccttaaacacggggtctgcaactagtcacaccca 359
QY 662 ccaatc-----agagagctcactaaatgctaaatcagggcaaaaagaggtaaag 712
DB 360 ccaatcaggtagtaaaagagagctcactaaatgctaaatcagggcaaaaagaggtagag 419
QY 713 caatagccaatcactatctgctgtagagcagcaggaaggaacagagattggtgataaa 772
DB 420 aaatagccaatcactatctgctgtagagcagcaggaaggaacagagattggtgataaa 479
QY 773 ctcaagcatcaacagcagcaacagcaaccccttggtgctccctccatgtagtggagc 832
DB 480 cccaagcatcagagccagcaacgctgccccttggtgctccctccatgtagtggagc 539
QY 833 tctgtttcactcattactcactatataatcaggaacagcagcactcttctggtcgtgt 892
DB 540 tctgtt-----ttccactcattataatcaggaacagcagcactcttctggtcgtgt 589
QY 893 ttctatgctcaagctgagcttctgttcgcacatccacacagctggttgcacccgtaca 952
DB 590 tgttaagcttgaagctgagcttctgttcgcacatccacacagctggttgcacccgtaca 649
QY 953 gaaccgctgctgacttccatccctttgagtcacagagtgctcactgctgctatcc 1012
DB 650 gactcgctgctgacttccatccctttgagtcacagagtgctcactgctgctatcc 708
QY 1013 agcaggttaccatctgacatcccgatccaggtgaaagcttgcatgttccctgacggc 1072
DB 709 agcaggttaccatctgacatcccgatccaggtgaaagcttgcatgttccctgacggc 768
QY 1073 taagtgcctgggttctcctaataagaaactgaaacactgctcaactggttctctc 1132
DB 769 taagtgcctgggttctcctaataagaaactgaaacactgctcaactggttctctc 828
QY 1133 tccatgacccaagcttccataatagactataaacctacacagatgcccagaattccatc 1192
DB 829 tccatgacccaagcttccataatagactataaacctacacagatgcccagaattccatc 888
QY 1193 ccttgatctgtaggagcgaaggaaccccaagctgaggaagtggtggttgcacacttg 1252
DB 889 ccttgatctgtaggagcgaaggaaccccaagctgaggaagtggtggttgcacacttg 948
QY 1253 ggaagtggcccaactgcatcttgtagcggcccaacacatcttgtagagctgtaggagca 1312
DB 949 ggaagtggcccaactgcatcttgtagcggcccaacacatcttgtagagctgtaggagca 1008
QY 1313 aggatccccagtaaca 1329
DB 1009 aggatccccagtaaca 1025

RESULT 5
AAA63826
ID AAA63826 standard; DNA; 2030 BP.
XX

```

AC   AAA63826;
XX   04-DEC-2000 (first entry)
XX   Nucleotide sequence of the MSRV-1 3' env and LTR regions.
DE   MSRV-1: pol region; long terminal repeat; LTR; RUS region; retrovirus;
KM   89.
XX   Multiple Sclerosis retrovirus 1.
OS   1..1629
FH   Location/Qualifiers
FT   CDS
FT   sig_peptide
FT   CAAT_signal
FT   CAAT_signal
FT   TATA_signal
FT   polyA_signal
FT   tag- f
XX   MO200047745-A1.
XX   17-AUG-2000.
XX   15-FEB-2000; 2000WO-1B00159.
XX   15-FEB-1999; 99EP-0420041.
XX   (INMR ) BIO MERIEUX.
XX   Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX   WPI: 2000-506097/45.
XX   P-PSDB: AAB08195.
XX   Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
XX   retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
XX   biological sample.
XX   Disclosure: Fig 2; 23pp; English.
XX   The present sequence represents the nucleotide sequence corresponding
XX   to the 3' env region and long terminal repeat sequences from clone
XX   CL6 of Multiple Sclerosis retrovirus (MSRV-1). The specification
XX   describes a long terminal repeat (LTR)-RUS region which encodes the
XX   expression of a MSRV-1 protein. This is unusual for LTRs, in
XX   particular in the RUS region. The sequence includes CAAT and TATA
XX   signals which are present in the U3 and R regions and are not directed
XX   towards the CDS indicated in the features table. Probes and antibodies
XX   to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
XX   are used to detect the presence of MSRV-1 retrovirus in a biological
XX   sample.
XX   Sequence 2030 BP: 574 A; 559 C; 387 G; 510 T; 0 other:
SQ

```

```

Query Match      58.3%; Score 775.4; DB 21; Length 2030;
Best Local Similarity 92.8%; Pred. NO. 1.1e-243;
Matches 812; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

```

QY   1 tcaaatcgaagagcttagactgtaacgcgcaaaagagggaacgtttattt 60
    |||||
DB   1140 tcaaatcgaagagcttagactgtaacgcgcaaaagagggaacgtttattt 1199
QY   61 agggagaaagagctgtgtgtatgtaacatctggaactcttctgtgaaagttaaaga 120
    |||||
DB   1200 agggagaaagagctgtgtgtatgtaacatctggaactcttctgtgaaagttaaaga 1259

```

```

QY   121 aattgagatcgaatatagttagagcagaggaccttcaaaacatgcacctggggcct 180
    |||||
DB   1260 aattcagagatcgaatataatgttagagcagaggacgttcaaaacacagcagctggggcct 1319
QY   181 cctcagccaatggatgagccctggacctctccctctttagagcctcagaagctataatc 240
    |||||
DB   1320 cctcagccaatggatgagccctggatcctccctctttagagcctcagaagctataatc 1379
QY   241 tttaactctcttggacccttcatctcaactctctttagtctgtctcttcagat 300
    |||||
DB   1380 gttactctcttggacccttcatctcaactctctttagtctgtctcttcagat 1439
QY   301 tgaagctgtaagctacaataatgtcttcaaaatggaaacccagatgagctcatgactaa 360
    |||||
DB   1440 tgaagctgtaagctacaagatgtcttcaaaatggaaacccagatgagctcatgactaa 1499
QY   361 aatcaccgtgtagccctcgagccgctcgcttagatagctctgtatgttaatgatgga 420
    |||||
DB   1500 gatccaccgtgtagccctcgagccgctcgcttagatagctctgtatgttaatgatgga 1559
QY   421 agtaccctcccgaggaaatctcaactgcaaacccctactacatcctaattcagtag 480
    |||||
DB   1560 agtaccctcccgaggaaatctcaactgcaaacccctactacatcctaattcagtag 1619
QY   481 aagcagttagagcaggtgtcagccaacctccccaacagtaactgggtttcctgttga 540
    |||||
DB   1620 aagcagttagagcaggtgtcagccaacctccccaacagtaactgggtttcctgttga 1679
QY   541 ggggtgactgtagagcagagtagtggatttccatgagctgactaagatccomaagcct 600
    |||||
DB   1680 ggggtgactgtagagcagagtagtggatttccatgagctgactaagatccomaagcct 1739
QY   601 aactggaaggtgagcagatccatcttcaaaatgggctgcaacttagctcacaccg 660
    |||||
DB   1740 aactggaaggtgagcagatccatcttcaaaatgggctgcaacttagctcacaccg 1799
QY   661 accaatcagagagctactaaatgtctaatcaggaacaaacagagaggttaagcaatagcc 720
    |||||
DB   1800 accaatcagagagctactaaatgtctaatcaggaacaaacagagaggttaagcaatagcc 1859
QY   721 aatcctatgtgctcgaagacagcagcggaaggaagatgtggaatataatcgaagca 780
    |||||
DB   1860 aatcctatgtgctcgaagacagcagcggaaggaagatgtggaatataatcgaagca 1919
QY   781 ttcaagccagcaacagcaaccccttgggtccctcccatgtgtagtggagctctgttt 840
    |||||
DB   1920 ttcaagccagcaacagcaaccccttgggtccctcccatgtgtagtggagctctgttt 1979
QY   841 cactctattcaactatataatcatgcaactgca 875
    |||||
DB   1980 cactctattcaactatataatcatgcaactgca 2014

```

```

RESULT 6
AAK77526
ID   AAK77526 standard; cDNA; 2946 BP.
XX
XX   AAK77526;
XX
XX   10-AUG-1999 (first entry)
XX
XX   Human secreted protein AJ172_2 cDNA.
XX
XX   Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
XX   bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
XX   cell proliferation; cell differentiation; suppressor; tumour inhibition;
XX   haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
XX   haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
XX   cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
XX   Homo sapiens.
XX

```

CC This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.

OY	421	agcaccctcccgaggaatcttaactgacacacccctactaactcaatccagtagg	480
Db	2475	aggcacccctctctgaagaatcctcagctgcacacccctcttaagcccaatccaagay	2533
OY	481	aagcagttagacaggttgcagccaacctccccacagtaacttgggtttctctgttga	540
Db	2535	aagcagttagagcggtgcgtcgccaacctccccacagcaacttagtttccgtttgaga	259
OY	541	gggtgagctgagaaacagagactatctgattctccttaagctcgactaagaatccnaagct	600
Db	2555	tggggagctgagaaacagagactagctggattctccttaagctcgactaagaatccctaagct	265
OY	601	ancitgggaaggtagccgagcttcactttaaacaatggggctgtgcacactagctcacaccg	660
Db	2655	agctggagaaggtgacacacatccacctttaaacaacagggcttgcacacttagctcaacctg	271
OY	661	accatccagaagagctctcaataatgctcaatcaggcaaaaacagagagttaagcaatagcc	720
Db	2715	accatccagaagagctctcaataatgctcaatcaggcaaaaacagagagttaagcaatagcc	277
OY	721	aatcatctatgctctgagagacagcgcggaagagacaagattggatataactcagca	780
Db	2775	aatcatctatgctctgagagacagcgcggaagagacaagattggatataactcagca	283
OY	781	ttcaagccagacaagacaacccctttgggtccctcccatgtgatgtgagctctgttt	840
Db	2835	ttcgcgcgcgacaacgcaacccctttgggtccctcccttggatgtgagagctctgttt	289
OY	841	cactcatcttctcctctatcaatcatgcaactgca	875
Db	2895	catgctatctcactctatcaatcttgcacactgca	2929
RESULT	7		
AAZ59468			
ID	AAZ59468	standard; cDNA; 2946 BP.	
XX	AAZ59468;		
AC			
XX	11-APR-2000 (first entry)		
DT			
XX			
DE	Human secreted protein Auj172_2 polynucleotide sequence.		
XX			
KM	Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;		
KM	placental pathology; metastasis inhibition; nutritional activity;		
KM	immune stimulator; haematopoiesis regulator; tissue growth;		
KM	tumour inhibitor; anti-inflammatory; clone Auj172_2; ATCC_98115;		
XX	gene therapy; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W09960020-A1.		
XX			
PD	25-NOV-1999.		
XX			
PF	17-MAY-1999; 99WO-US10915.		
XX			
PR	18-MAY-1998; 98US-0080478.		
PR	20-OCT-1998; 98US-0175928.		
XX			
PA	(GEMV) GENETICS INST INC.		
XX			
PI	Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;		
PI	Meiberg D, Mi S, Treacy M;		
DR	WP1: 2000-116311/10.		
XX			
DR	P-PSDB; AAY67313.		
XX			
PT	New polynucleotides encoding secreted cDNA libraries, used to develop		
XX	products for the diagnosis and treatment of neoplastic disease		
PS	Claim 14; Page 107-108; 149pp. English.		

This is the human secreted protein A172-2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating activity (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide sequences are also stated to be useful for gene therapy.

Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other:

Query Match 55.1%; Score 731.8; DB 21; Length 2946;

Best Local Similarity 90.7%; Pred. No. 2,9e-229;

Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 tcaaaatcgaagagcttagacttgcataccgcaaaagagggggaacctgttatctt 60
DB 2067 tcaaaatcgaagagcttagacttgcataccgctgaagagggggaacctgttatctt 2126
QY 61 aggggaagaagctgtgtgtatgttaataatctggaactcttactggaagaatgaaga 120
DB 2127 aggggaagaagctgtgtgtatgttaataatctggaactcttactggaagaatgaaga 2186
QY 121 aattgaagatcgaatataatctagaagcagagagacctcaaaacctgcagctggagcct 180
DB 2187 aattcgaagatcgaatataatctagaagcagagagagctctgaaacctgcagctggagcct 2246
QY 181 cctcgaagatcgaatgtgccttgcagctcctcctctttagagcctctagagctataatatt 240
DB 2247 cctcgaagatcgaatgtgccttgcagctcctcctctttagagcctctagagctataatatt 2306
QY 241 tttaactctcttgcagcctgtatcttcaactctcctgttgaattgtctcttcagagat 300
DB 2307 gctactctcttgcagcctgtatcttcaactctcctgttgaattgtctcttcagagat 2366
QY 301 tgaagctgtgaagcctacaaatagttctcaaatggaaccagagatgcagctcatgactaa 360
DB 2367 cgaagctgtgaagcctacaaatagttctcaaatggaaccagagatgcagctcatgactaa 2414
QY 361 aatcctacgtgagccctcgcagcggcctgtcagactatgctctgaattgaatgaactga 420
DB 2415 gatctacccgagaccctcgcagcggcctgtcagactatgctctgaattgaatgaactaa 2474
QY 421 agtcccccctccgaggaatctcaactgcagaaaccccttacttaactcaattagtgagg 480
DB 2475 agtcccccctccgaggaatctcaactgcagaaaccccttacttaactcaattagtgagg 2534
QY 481 aagcagcttagagcagcttctcagccaacctccccaacagactcttggtttctcgttgaaga 540
DB 2535 aagcagcttagagcagcttctcagccaacctccccaacagactcttggtttctcgttgaaga 2594
QY 541 gggtgagctgagagacagagactctggaattcctcctagcctgactaagaatcccaagcct 600
DB 2595 tggggtgagctgagagacagagactctggaattcctcctagcctgactaagaatcccaagcct 2654
QY 601 ancggtgagagctgagacagactcttcaaacatggggctgcagacttagctacacccg 660
DB 2655 agcttgggaagctgagacagactcttcaaacatggggctgcagacttagctacacccg 2714
QY 661 accaatcgaagagctcactataatctgaatcagcaaaacagagagctgaagcactatagcc 720
DB 2715 accaatcgaagagctcactataatctgaatcagcaaaacagagagctgaagcactatagcc 2774

QY 721 aatcatctatgcttgaagcagcagcggaagacaaagatctggatataaactcagga 780
DB 2775 aatcatctatgcttgaagcagcagcggaagacaaagatctggatataaactcagga 2834
QY 781 ttcaatcgaagacagacagaccccttgggtccctccatctgatatggagctctgttt 840
DB 2835 ttcaatcgaagacagacagaccccttgggtccctccatctgatatggagctctgttt 2894
QY 841 cactcatcactatcaatcaatcatgcaactga 875
DB 2895 catgctattcactatcaatcaatcttgcactga 2929

RESULT

AAFS5630 ID AAF55630 standard; DNA: 2781 BP.

AC AAF55630;

DT 29-MAY-2001 (first entry)

DE Nucleotide sequence of a human endogenous retrovirus envelope protein.

KW Envelope protein; HERV; syncytia formation; placental development;

KW syncytia; cancer; cell adhesion; ss.

OS Human endogenous retrovirus.

FH Key Location/Qualifiers

FT CDS 762..2378

FT /"tag" a

FT /product= "envelope protein"

PN MO200116171-A1.

PD 08-MAR-2001.

PF 01-SEP-2000; 2000WO-FR02429.

PR 01-SEP-1999; 99FR-0011141.

PR 15-SEP-1999; 99FR-0011793.

PA (INMR) BIO MERIEUX.

PA (INRM) INST NAT SANTE & RECH MEDICALE.

PI Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;

DR WPI: 2001-226676/23.

DR P-PSDB: AAB67652.

PT Detecting expression of human endogenous retrovirus envelope protein in

XX cells of a tissue or culture, from its ability to induce syncytia -

PS Disclosure; Page 44-45; 57pp; French.

XX The present sequence encodes a human endogenous retrovirus envelope

CC protein. The specification describes a method for detecting expression

CC of an envelope protein from a human endogenous retrovirus (HERV), in

CC cells, of a tissue or culture. The method comprises detecting syncytia

CC formation due to the fusogenic properties of the envelope protein.

CC Envelope polypeptides and polynucleotides are used to produce

CC therapeutic or prophylactic compositions, particularly for treatment of

CC cancer, to correct defects in placental development (or other natural

CC formation of other types of syncytia), and to promote adhesion of cells

CC in grafts or cellular repair processes. Expression of sequences

CC antisense to the polynucleotide are used to prevent formation of

CC syncytia.

XX Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;

Query Match 54.9%; Score 730.2; DB 22; Length 2781;

Best Local Similarity 90.6%; Pred. No. 9,5e-229;


```

Oy 241 tttaactctcttgagacccttatcttcaactctctgttaagtctgtctcttccagaat 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1415 gctactctcttgagacccttatcttcaactctctgttactctgtctctccagaat 1474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 301 tgaagctgttaagcctacaatagtcttcaatgtgaaccagatgcagctcatgactaa 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1475 cgaagctgttaaaacta-----caaatgtgaaccagatgcagctcatgactaa 1522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 361 aactaaccgtgagaccct-ggaccggcctctgtaagctatgctctgagttaatgactg 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1523 gatctaaccgagaccctcgagaccgctctgtaaccagatctgagttaatgacta 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 420 aagtaaccctcccgaggaatactcaactgcacaaccctcatcactc----- 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1583 aaggcaaccctctgaggaatactcaagctgcacaaccctcatcactcagccaa 1642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 470 -----aactcagtgaggaagcagcttagagcagctgtaagcgaaccctcccaagctactg 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1643 cagggaagcagcttagagagcggtctgctggccaaccctcccaagcagcttagg 1702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 525 ggtctctctgtgagaggtgagctgagagag-----acagagtaagctgga--tttctag 578
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1703 gtttctctgttggagatgggggagctgagagcagggagttagctgtggttctctag 1762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 579 ctgactaagaatcccaagcctantctgggaaggtgacg-----catccatct 626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1763 ctgacctaagaatcccaagcctantctgggaaggtgacg-----catccatct 1822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 627 ttaacaatgggg-----ctgcaactagctacacaccc--gaccaatcagagagctacata 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1823 taabaacagggggctcttgcaactttagctcacaccttgcaactcagagagctacata 1882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 681 aaatg-ctaactcagcaaaacacagaggttaagcaatagccaatcatctatctgctga-g 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1883 aaatgctaactcagcaaaacacagaggttaagcaatagccaatcatctatctgctgag 1942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 739 agcacaagcggaagagacagaagatgggataataactcagcagatccaagccagc--aacagc 797
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1943 agcacaagcggaagagacagaagatgggataataatacccgacatcagccagcaagc 2002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 798 aaccctcttggtccctccatctgataggagctcgt-----ttcaactta 847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2003 aaccgctcttggtccctccatctgataggagctcgt-----ttcaactta 2062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 848 ttcaactatataatcagtaactg--caactctctgctgctggttttatagtctcaa 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2063 ttcaactatataatcagtaactgtaactctctctgctgagtggttttataagctcta 2122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 906 gctgagctcttgctgcacacacacgctgctgtaacacgctca-----cagaaccgctgc 962
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2123 gctgagctcttgctgcacacacacgctgctgtaacacgctgtaacacgctgc 2182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 963 tgaactcaccctcttgatccagcaaggtgtccacgtgtctctgtaacccagaggtac 1022
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2183 tggctcaccctcttgatccagcaaggtgtgtctgctgcatctctgtaacccagaggtac 2242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1023 ccaatgcaactcccgatcaggtctaaaggtctgcaatgtctcctgcaatgtaagctgac 1082
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2243 ccaatgcaactcccgatcaggtctaaaggtctgcaatgtctcctgcaatgtaagctgac 2302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1083 ggttctgcttaataagaaactgtaacactgctggtctgcaatgtctctctccatgacc 1142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2303 ggttctgcttaataagaaactgtaacactgctggtctgcaatgtctctctccatgacc 2362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1143 agcgctcttaataag--aactataacacacacgctgcaatgtctcctgta 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2363 atggtctcttaataagaaactataacacacacgctgcaatgtctcctgta 2422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1201 --tctgtgagcgcaagaaccccaagctcagagaangtgagcttgccaacatcttggaagt 1258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2423 atctcgtagagcgcaagaaccccaagctcagagaangtgagcttgccaacatcttggaagc 2482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1259 ggcaccaactgcatcttgtagcggtcccaacacatcttgtagagcttgtaggaagaagatc 1318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 2483 ggcggctgcacactctggaagcggtctgcccacacatcttgagagctctgtagcaagaagcc 2542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1319 cccagtagta 1329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2543 ccccggtataca 2553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 bp.
XX
AC AAK25665:
XX
DT 21-MAY-1999 (first entry)
XX
DE Complete human endogenous retrovirus W genome.
XX
KW Clone: human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
OS Human endogenous retrovirus.
XX
PN W09902696-A1.
XX
PD 21-JAN-1999.
XX
PE 06-JUL-1998; 98WO-FR01442.
XX
PR 07-JUL-1997; 97FR-0008815.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Beseme F, Blond JL, Bouton O, Maillet F, Mandrand B;
DR WPI: 1999-120897/10.
XX
PT New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
PS Claim 1; Page 71-74; 106pp: French.
XX
CC This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. Multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin-dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.
XX
SQ Sequence 7582 bp; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 54.4%; Score 722.6; DB 20; Length 7582;
Best Local Similarity 88.5%; Pred. No. 5.5e-226;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;

```

Db 6900 cctcagccratgatgctccgagattctccctctcttagagcccttagcagctataatt 6959
Qy 241 ttatccctcttggagccctgtatcttcaactccttgttaagtgtctctccgaat 300
Db 6960 gctacccctcttggagccctgtatcttcaactccttgttaagtgtctctccgaat 7019
Qy 301 tgaagctgtaagaactcaatagttcttcaatggaaccccgatgcagtcagtactaa 360
Db 7020 cgaagctgttaacta-----caaatggagcccaagatgcagtcgaactaa 7067
Qy 361 aatctacgttggacccctggaccgctctgtagactatgctctgagttaagacttga 420
Db 7068 gatctacggagaccctggaccgctctgtagcagcagctctgagttaagacttga 7127
Qy 421 agtcacccctccggaggaatctcaactgacacacccctactcaactcaatcagtagg 480
Db 7128 aggcacccctccggaggaatctcaactgacacacccctactcaactcaatcagtagg 7187
Qy 481 aagcagctagagagctgtctcagcgaacctcccaacagtaactgtgggtttctcgttgaga 540
Db 7188 aagcagctagagagctgtctcagcgaacctcccaacagtaactgtgggtttctcgttgaga 7247
Qy 541 gggtagctgagagagagactagctgagattctcctagagctgactaagaalccnaagcct 600
Db 7248 tggggagctagagagagactagctgagattctcctagagctgagtagaagatcgttagagcct 7307
Qy 601 anctggagaggtgtagcgcctccatctttaaactgtgggctgtagacttagctacacccg 660
Db 7308 agstggagaggtgtagcgcctccatctttaaactgtgggctgtagacttagctacacccg 7367
Qy 661 accaactag 720
Db 7368 accaactag 7427
Qy 721 aatcatctattgctctgtag 780
Db 7428 aatcatctattgctctgtag 7487
Qy 781 ttcaagcag 840
Db 7488 ttcaagcag 7547
Qy 841 cactctattctactctatcaatcaatcagtagcagtagcagtagcagtagcagtagcagtagc 875
Db 7548 catgctattctactctatcaatcaatcagtagcagtagcagtagcagtagcagtagcagtagc 7582

SQL* 11
ID AA59215 standard; DNA; 7582 BP.
AC AA59215;
DT 07-NOV-2000 (first entry)
DE Human endogenous retrovirus W (HERV-W) sequence.
XX
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT LTR 1..120
FT LTR /tag= a
FT LTR /note= "R of 5' LTR"
FT LTR 121..575
FT LTR /tag= b
FT LTR /note= "U5 of 5' LTR"
FT primer_bind 579..596
FT CDS /tag= c
FT CDS 5581..7194

FT /tag= d
FT /note= "ORF1 env538"
FT CDS 7039..7194
FT /tag= e
FT CDS /note= "ORF2 52 AA"
FT 7112..7255
FT /tag= f
FT /note= "ORF3 48 AA"
FT misc_feature 7244..7254
FT /tag= g
FT /note= "polypurine tract"
FT LTR 7256..7582
FT /tag= h
FT /note= "U3-R of 3' LTR"
FT polyA_signal 7563..7569
FT /tag= i
PN W0200043521-A2.
PD 27-JUL-2000.
PF 21-JAN-2000; 2000MO-FR00144.
PR 21-JAN-1999; 99PR-0000888.
PA (INMR) BIO MERIEUX.
PI Paranhos-Baccala G, Mallet F, Voisset C;
DR WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
XX diagnosis of autoimmune disease and complications of pregnancy,
XX contains at least part of the gag gene
XX
XX
XX Disclosure; Page 49-52; 53pp; French.
PS
PS
XX
XX The present sequence represents an endogenous retrovirus, which is
XX associated with an autoimmune disease, and is integrated into the human
XX genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
XX HERV-W retrovirus is associated with autoimmune disease, failure of
XX pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
XX proteins derived from it, are useful for diagnosis of autoimmune
XX disease (specifically multiple sclerosis) and for monitoring pregnancy.
XX The nucleic acid fragments may also be used for in situ labelling of
XX isolated chromosomes, while the transcription product can be used to
XX study or monitor T cell proliferation in vitro.
XX
SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;
XX
Query Match 54.4%; Score 722.6; DB 21; Length 7582;
Best Local Similarity 88.5%; Pred. No. 5.5e-726;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;
Qy 1 tcaaatcgaaagagcttagactgtctaacccgcaaaagagggggaacctgtatttt 60
Db 6720 tcaaatcgaaagagcttagactgtctaacccgctgaaggggggaacctgtatttt 6779
Qy 61 aggggaagatgctgttagatgttaatcaatctggaatcaatcagtagaagaatlaaga 120
Db 6780 aggggaagatgctgttagatgttaatcaatctggaatcaatcagtagaagaatlaaga 6839
Qy 121 aatttgatcgaatataatgttag 180
Db 6840 aatttgatcgaatataatgttag 6899
Qy 181 cctcagcgaatgag 240
Db 6900 cctcagcgaatgag 6959
Qy 241 ttactcctcttggagccctgtatcttcaactccttgttaagtgtctctccgaat 300

QY 1 tcaaaatcgaagagctttagacttgtaacccgcaaaagagggaacctgttatttt 60
 |||||
 Db 1902 tcgaaatcgaaagagctttagacttgtaacccgctgagagagggaacctgttatttt 1961
 |||||
 QY 61 aggggaagaatgctgttagtattgtaacaaatcgtgaatcattactgagaagttaaaga 120
 |||||
 Db 1962 aggggaagaatgctgttagtattgtaacaaatcgtgaatcattactgagaagttaaaga 2021
 |||||
 QY 121 aatttgaagtctgaataatgtagagcagagggacttcaaaaacatgcacccctggggcct 180
 |||||
 Db 2022 aatttcaagatcgaaatacgaatacgaagagcttcgaataacatcgacccctggggcct 2081
 |||||
 QY 181 cctcagccaatgtagatgccttgagacttccctctttagagacccttagcagctataatt 240
 |||||
 Db 2082 cctcagccaatgtagatgccttgagacttccctctttagagacccttagcagctataatt 2141
 |||||
 QY 241 tttaactcctcttggaacctgtatcttcaactcctctgttaagtgtctctccagaat 300
 |||||
 Db 2142 gctaccctcttggaacctgtatcttgaacctctgttaacttgcctctccagaat 2201
 |||||
 Db 301 tgaagctgtaaaactaataatgattcttcaaatggaaccccaatgcagtcacatgactaa 360
 |||||
 Db 2202 cgaagctgtaaaacta-----caaatggaaccccaatgcagtcacatgactaa 2249
 |||||
 QY 361 aatctacgttggaaccttggaacctgagcctgtgtagactatgctcgaatgtaataatgta 420
 |||||
 Db 2250 gatctacgcgagaccccttggaacctgagcctgtgtagactatgctcgaatgtaataatgta 2309
 |||||
 QY 421 agtcacccctccgaggaatctcaacttgcaacacccctactaactccaatcagtaag 480
 |||||
 Db 2310 aggcacccctccgaggaatctcaacttgcaacacccctactaactccaatcagtaag 2369
 |||||
 QY 481 aagcaattgagagagtgtagcagcaacccctcccaacagtaactggtttccctgttga 540
 |||||
 Db 2370 aagcaattgagagagtgtagcagcaacccctcccaacagtaactggtttccctgttga 2429
 |||||
 QY 541 ggttgactagagagagagtagctgagatttcttagagctgactaagaatccnaagcct 600
 |||||
 Db 2430 ggttgactagagagagtagctgagatttcttagagctgactaagaatccnaagcct 2489
 |||||
 QY 601 anctgggaaggtgacgcgacatcactttaaacaatgggcttgcaacttagtcacacccg 660
 |||||
 Db 2490 anctgggaaggtgacgcgacatcactttaaacaatgggcttgcaacttagtcacacccg 2549
 |||||
 QY 661 accaatcagagagtagctactaataatgttaatacagcaaaaacagaggtaaagcaatagcc 720
 |||||
 Db 2550 accaatcagagagtagctactaataatgttaatacagcaaaaacagaggtaaagcaatagcc 2609
 |||||
 Db 721 aatcatctatgtcctgagagcagcggaagagagaagattgggatatataaactcagca 780
 |||||
 Db 2610 aatcatctatgtcctgagagcagcggaagagagaagattgggatatataaactcagca 2669
 |||||
 QY 781 ttcaagccagcaaacaccccttbggtlccctcccaatgtatgagagctctgttt 840
 |||||
 Db 2670 ttcaagccagcaaacaccccttbggtlccctcccaatgtatgagagctctgttt 2729
 |||||
 QY 841 caactatttcaactatttaataatcagtaactgca 875
 |||||
 Db 2730 caactatttcaactatttaataatcagtaactgca 2764
 |||||

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:52:31 ; Search time 4429.18 Seconds
(Without alignments)
6279.123 Million cell updates/sec

Title: US-09-319-156A-12
Perfect score: 1329
Sequence: 1 tcaaatcgagagacttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_htg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_cm: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_htg_hum: 31: em_htg_inv: 32: em_htg_other: 33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	1326	99.8	1329	6	AX001030	AX001030 Sequence
2	1316.4	99.1	163803	9	AC093531	AC093531 Homo sapi
3	1035.2	77.9	56093	6	AX329572	AX329572 Sequence
4	1035.2	77.9	56093	9	HSAC000064	AC000064 Human BAC
5	1035.2	77.9	149134	9	AC007566	AC007566 Homo sapi
6	1023.2	77.0	10499	6	AX007980	AX007980 Sequence
7	1006.6	75.7	188919	2	AC009727	AC009727 Homo sapi
8	989.6	74.5	161571	2	AC092708	AC092708 Homo sapi
9	989.6	74.5	166700	9	AC013759	AC013759 Homo sapi
10	989.6	74.5	186911	2	AP001488	AP001488 Homo sapi
11	985.6	74.2	118241	9	AP001599	AP001599 Homo sapi
12	985.6	74.2	340000	9	AP001697	AP001697 Homo sapi
13	984	74.0	91097	9	AP001600	AP001600 Homo sapi
14	982.6	73.9	172721	2	AC008121	AC008121 Homo sapi
15	979.2	73.7	145785	2	AC016173	AC016173 Homo sapi
16	968.8	72.9	157711	9	AC055738	AC055738 Homo sapi
17	968.8	72.9	207945	9	CNS01DS	AC009946 Homo sapi
18	951.4	71.6	169072	9	AC009946	AC009946 Homo sapi
19	951.2	71.6	162522	9	AC093839	AC093839 Homo sapi
20	948.2	71.3	155925	9	AC096536	AC096536 Homo sapi
21	912.8	68.7	183680	9	AC098859	AC098859 Homo sapi
22	912.8	68.7	221525	2	AC021652	AC021652 Homo sapi
23	876.4	65.9	174019	9	AP001538	AP001538 Homo sapi
24	876.4	65.9	340000	9	AP001674	AP001674 Homo sapi
25	870.4	65.5	83412	9	AC092843	AC092843 Homo sapi
26	859	64.6	169462	2	AC024033	AC024033 Homo sapi
27	851.2	64.0	251124	9	HUAE000660	AB000660 Homo sapi
28	833.8	62.7	140756	9	ALI39038	ALI39038 Human DNA
29	833.8	62.7	183499	2	AL607153	AL607153 Homo sapi
30	833	62.7	839	9	AF127226	AF127226 Homo sapi
31	824.8	62.1	114455	9	AC093742	AC093742 Homo sapi
32	816.6	61.4	187399	9	AC067942	AC067942 Homo sapi
33	811.4	61.1	153104	2	AC008564	AC008564 Homo sapi
34	809.8	60.9	114972	9	AC016617	AC016617 Homo sapi
35	808.2	60.8	181016	2	AC073129	AC073129 Homo sapi
36	801.6	60.3	156821	9	AC008448	AC008448 Homo sapi
37	798.2	60.1	88328	9	AL357874	AL357874 Human DNA
38	789.6	59.4	172182	2	AC091714	AC091714 Homo sapi
39	777.6	58.5	188038	2	AC092510	AC092510 Papio cyn
40	768	57.8	164310	9	AC016156	AC016156 Homo sapi
41	768	57.8	182224	2	AC093116	AC093116 Homo sapi
42	766	57.6	117124	2	AC110614	AC110614 Homo sapi
43	762.4	57.4	123253	9	AC078963	AC078963 Homo sapi
44	762.4	57.4	138003	2	AC018346	AC018346 Homo sapi
45	762.4	57.4	158059	2	AC079947	AC079947 Homo sapi

ALIGNMENTS

RESULT	1	1329 bp	DNA	linear	PAT 10-MAR-2000
AX001030	AX001030	Sequence 12 from Patent WO9902666.			
LOCUS	AX001030				
DEFINITION	Sequence 12 from Patent WO9902666.				
ACCESSION	AX001030				
VERSION	AX001030.1	GI:7241266			
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 1329)				
AUTHORS	Ott,C. and Bedin,F.				
TITLE	RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES				
JOURNAL	PATENT: WO 9902666-A 12 21-JAN-1999.				
FEATURES	BIO MERIEUX (FR); OTT CATHERINE (FR)				
SOURCE	Location/Qualifiers				
	1..1329				
	/organism="unidentified"				
	/db_xref="taxon:32644"				

BASE COUNT	346 a	352 c	288 g	340 t	3 others
------------	-------	-------	-------	-------	----------

Query Match 99.8%; Score 1326; DB 6; Length 1329;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 tcaaatcgaagagcttagacttgcataccgcaaaagaggaggaactgttattt 60
D 1 tcaaatcgaagagcttagacttgcataccgcaaaagaggaggaactgttattt 60
QY 61 aggggaagaagctgttgaatgttataatcaatctcgaaatcaatctactgaaagttaaga 120
D 61 aggggaagaagctgttgaatgttataatcaatctcgaaatcaatctactgaaagttaaga 120
QY 121 aatttgagatcgaaatataatgttaagacagagagacttcaaaacactgcacacctggagcct 180
D 121 aatttgagatcgaaatataatgttaagacagagagacttcaaaacactgcacacctggagcct 180
QY 181 cctaagcaatcgatgccttgccttcccttcttaagaccttgaagactataatatt 240
D 181 cctaagcaatcgatgccttgccttcccttcttaagaccttgaagactataatatt 240
QY 241 tttaactcttcttgaaaccttatacttaacttcttcttgtaagtgttctcttcagagat 300
D 241 tttaactcttcttgaaaccttatacttaacttcttcttgtaagtgttctcttcagagat 300
QY 301 tgaagctgttaagcttacaatagttcttcaaaatggaacccacagatgcagttcactaa 360
D 301 tgaagctgttaagcttacaatagttcttcaaaatggaacccacagatgcagttcactaa 360
QY 361 aattaccgctgagacccctggagccgctcctgaagactaagcttgaatgttaatgacattga 420
D 361 aattaccgctgagacccctggagccgctcctgaagactaagcttgaatgttaatgacattga 420
QY 421 agtcaacccctccgaggaatctcaacttgcacaaccccttactaacttcaatctcaatga 480
D 421 agtcaacccctccgaggaatctcaacttgcacaaccccttactaacttcaatctcaatga 480
QY 481 aagcaatgaagcagttgttcagcacaacctcccaacagttacttgggtttccctgttga 540
D 481 aagcaatgaagcagttgttcagcacaacctcccaacagttacttgggtttccctgttga 540
QY 541 ggggtgagctgagagagagactagttcttcttgaagcttactaagaatcccaagact 600
D 541 ggggtgagctgagagagagactagttcttcttgaagcttactaagaatcccaagact 600
QY 601 aactcgaagagtgagacgcatcttcaaacatggggtctgcacacttgcacacccg 660
D 601 aactcgaagagtgagacgcatcttcaaacatggggtctgcacacttgcacacccg 660
QY 661 accaatcgaagagctcactaaatgtctaaatcgaagaaacagagagtgaaagcaatagcc 720
D 661 accaatcgaagagctcactaaatgtctaaatcgaagaaacagagagtgaaagcaatagcc 720
QY 721 aatatctatctgcttgaagacagagcgaggaaggaagatctgggataataacacagga 780
D 721 aatatctatctgcttgaagacagagcgaggaaggaagatctgggataataacacagga 780
QY 781 ttcaagcgaagcaaacagaccccttgggtccctcccatctgtatggagagctgttctt 840
D 781 ttcaagcgaagcaaacagaccccttgggtccctcccatctgtatggagagctgttctt 840
QY 841 caactatcttactatcaataatgaatgcacacttcttgcgttcttcttcttctt 900
D 841 caactatcttactatcaataatgaatgcacacttcttgcgttcttcttcttcttctt 900
QY 901 ctcaagctgagcttcttgcacacacacacacacacacacacacacacacacacacacac 960
D 901 ctcaagctgagcttcttgcacacacacacacacacacacacacacacacacacacacac 960
QY 961 gcttacttcaatcccttggatccagcagagtgctcactgttcttcttcttcttcttctt 1020
D 961 gcttacttcaatcccttggatccagcagagtgctcactgttcttcttcttcttcttctt 1020

```

```

QY 1021 accacttgcacaccccgatccagctaaaggttgcatttcttcttcttcttcttcttctt 1080
D 1021 accacttgcacaccccgatccagctaaaggttgcatttcttcttcttcttcttcttctt 1080
QY 1081 tgggttgccttaataagacacacacacacacacacacacacacacacacacacacacac 1140
D 1081 tgggttgccttaataagacacacacacacacacacacacacacacacacacacacacac 1140
QY 1141 ccaggtcttcaatagactaactcaactcaactcaactcaactcaactcaactcaactcaact 1200
D 1141 ccaggtcttcaatagactaactcaactcaactcaactcaactcaactcaactcaactcaact 1200
QY 1201 tctgtgagcccaagaccccgatccagagagagtgaggttgcacacacacacacacacac 1260
D 1201 tctgtgagcccaagaccccgatccagagagagtgaggttgcacacacacacacacacac 1260
QY 1261 ccaactgcacatttgcgtgagcgcacacacacacacacacacacacacacacacacacac 1320
D 1261 ccaactgcacatttgcgtgagcgcacacacacacacacacacacacacacacacacacac 1320
QY 1321 ccagtaaca 1329
D 1321 ccagtaaca 1329
QY 1321 ccagtaaca 1329
D 1321 ccagtaaca 1329

RESULT 2
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS AC093531
DEFINITION Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
ACCESSION AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Nov 16, 2001 this sequence version replaced gi:15383820.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence:
Estimated Total Number of Errors is 0.3.
FEATURES
source
1. 163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"
BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN
Query Match 99.18; Score 1316.4; DB 9; Length 163803;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1320; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 tcaaatcgaagagcttagacttgcataccgcaaaagaggaggaactgttattt 60

```

```
Db 117604 TTTAAATTAAGAGCTTTTGACCTTCAACCGCCAAAGAGGGGGAACCTGTTATTTT 117545
Qy 61 aggggagaagatctgttaataatgaatcaatctggaatcattacgagaagtaaga 120
Db 117544 AGGGGAGAGATGCTGTAGATGTTATCAATGTGAATCAATTAAGGAAAGTTAAAG 117485
Qy 121 aatttagatcgataataatgtagacagagagacccctcaaacactgcaccgggacct 180
Db 117484 AATTGAGATCGAATTAATGTTAGACAGACAGACCTTCAAAACACTGCAACCGGGGCT 117425
Qy 181 cctcagcaatgtagagccctggaactcctcctcttaggaaccttagcagctaaatatt 240
Db 117424 CCTCAGCAGCAATGAGATGCTGAGACCTGCTCTTAGAGACCTTAGACGCTAATATAT 117365
Qy 241 tttaacctctcttgagaccctgtatctcaactcctctgttaagttgtctctccgaat 300
Db 117364 TTTACTCTCTTGGAGCCCTGTAATCTTAACTTGTAACTTGTCTCTTCCAGAAAT 117305
Qy 301 tgaagctgtaagctacaataatgctctcaaatggaacccagatgcaatgcaatgataa 360
Db 117304 TGAAGCTGTAAGCTCAATAATGTTCTTCAATGGAACCCAGATGCAATGCAATGACTAA 117245
Qy 361 aatctacgttgagaccctggaaccgctgctagactatgctctgataatgataatga 420
Db 117244 AATCTACGCTGAGACCCCTGGAGCGGCTGCTAGACTATGCTGTGATGTTAATGACATGTA 117185
Qy 421 agtaccctcccggaagaatctcaactgcaaacccctactactactcaatcgaatgag 480
Db 117184 AGTACCCCTCCCGAGGAATGCAACTGCAACCCCTAATCACTTCAATTCACATTAAGG 117125
Qy 481 aagcagctagaagctgtgcaagcaacctcccaacagtaactggtgttctcgttgaga 540
Db 117124 AAGCACTTAGAGAGAGTTGTACAGCAACCTCCCAACAGTACTTGGGTTTTCGTTGAGA 117065
Qy 541 ggggtgactggaagacagagactgtagattctcagagctgaagaaatccnaagct 600
Db 117064 GGGTGACGTGAGAGACAGGACTGTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCT 117005
Qy 601 anctggaaggtgagacgactcctcttaaaagtgagggtctgaaattgctcacacgg 660
Db 117004 AGCTGGGAAGGTAGCGCATCCATCTTTAAACATGGGGGCTGCAACTTACTACACCGG 116945
Qy 661 accaatcagagagctactaataatgctaaatcaggaacaaacagaggtlaaagaatagc 720
Db 116944 ACCAATCAGAGAGCTCACTAATAATGCTAATACAGCAAAAAACAGAGAGTAAGCAATAGCC 116885
Qy 721 aatcatctatctgctgagagacagcgaggaaggaagaatgtagatataaactcagga 780
Db 116884 AATCATCTATTGGCTTAGAGACAGCGGGAAGGAAAGGATTTGGATATTAACCCAGGCA 116825
Qy 781 ttcaagcagcaagaacaccccttggtccctcccaatgataagggagctctgttt 840
Db 116824 TTCAAGCCAGCAACACACCCCTTGGGTCCCTCCCTCATTTGATGGAGGCTGTGTTT 116765
Qy 841 cactcatctactctataatcaatcagcaactgcaactctctgtgcgtgttttttaag 900
Db 116764 CACTCATTTCACTATTAAATCTTGAACCTGCAACTCTCTGTGCTGCTTTTATATG 116705
Qy 901 ctaagcgtgagcttctgttcgcacatcaacatgctgtttgcaacgttcaagaaccgt 960
Db 116704 CTCACACTGAGCTTTTGTTCGCATCACACACGCTGTGTCACACGCTACACACCGCT 116645
Qy 961 gctgacttcaatccctttgtagatccagcaaggtgcaatgctgtctcagcagaggt 1020
Db 116644 GCTGACTTCAATCCCTTTGATCCAGACAGAGTGTCCATGCTCTCGATCCAGCGAGGT 116585
Qy 1021 acccatgccaactccagatcagcgtlaaaggctgccaatgttctcgtcagctgaatgac 1080
Db 116584 ACCCATGCGCACTCCGATAGGCTAAAGGCTTGAATGCTTCTGTGATGGCTAAAGTGC 116525
Qy 1081 tgggttctcctaataagaaactgtaacactgctggttccatggtctctccatgac 1140
Db 1140 tgggttctcctaataagaaactgtaacactgctggttccatggtctctccatgac 1140
```

```
Db 116524 TGGGTTTGCTTAATAGAACGTGAACACTGTGCTACGTGCTCTTCCATGAC 116465
Qy 1141 ccaagctcttaataagactataaactcaccgacatgagcccaagatctcattcttgta 1200
Db 116464 CCAGGCTTCTTAATAGAGCTAATACACTCACCGCATGGCCCAAGATTCCATTCCTGGTA 116405
Qy 1201 tctgttagcccaagaaccccaagctcagagaanaatgaggtctgacacatttggaatg 1260
Db 116404 TCTGTAGCGCAAGAACCCAGGTGAGAGACTGAGCGCTGCGACCATTTGGGAAGTGG 116345
Qy 1261 cccaactgcaatttgtagcgcccaaccacactcttggaagctgtaggaagatcc 1320
Db 116344 CCCACTGCCATTTTGGTAGCGGCCACACCATCTTGGAGCTGTGGAGCAAGATCCC 116285
Qy 1321 ccagtaaca 1329
Db 116284 CCAGTAAACA 116276
```

```
RESULT 3
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS AX329572 Sequence 81 from Patent WO0194629.
DEFINITION AX329572
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horrigan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
FEATURES
source Avalon Pharmaceuticals (US)
location/Qualifiers
1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN
```

Query Match 77.9%; Score 1035.2; DB 6; Length 56093;
Best Local Similarity 90.1%; Pred. No. 6.1e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

```
Qy 1 tcaaatcgaagagcttagacttgctaaacgcgaagaagggggaacctgtatttt 60
Db 37018 TCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAAGGGGGAACCTGTTATTTT 37077
Qy 61 aggggaagaatgctgtagtagtaatcaactggaatcattactgagaagtaaga 120
Db 37078 AGGGGAAGAAATGCTGTATATATGTTAATCAATCCGAATGCTCACTGAGAAAGTTAAAGA 37137
Qy 121 aatttagatcgataataatgtagagcagagagccttcaaaaacatgcaccctgggacct 180
Db 37138 AATTTCGAGATCGAATACAAAGTAGACAGAGAGGCTTTCGAACACTGAGACCCGTGGGCTT 37197
Qy 181 cctcagcaatgtagatccctggaactcctcctcttaggaaccttagcaactaataat 240
Db 37198 CTCACAGCAATGAGATGCTCGGATTTCTCCCTCTTAAAGACCTTAGACACTAATATAT 37257
Qy 241 tttaactctcttggaacctgtatcttcaactctcctgttaagtgtgctctccgaat 300
Db 37258 GCTACTCTCTCTTGGAGCCCTGTAATCTTAACTCTTGTAACTTGTCTCTTCCAGAAAT 37317
Qy 301 tgaagctgtaagaactacaataatgcttcaaatggaaccccaagatgagctgaatgata 360
Db 37318 CGAAGCTGTAAACTA-----CNAATGAGAGCCCAAGATGCACTCAAGACTAA 37365
```

```

Oy 361 aatcaccgtagaccctgagccgctgctagactatgctcgtatgtaagcaltga 420
    |||||||
Db 37366 GATCTACCGCAGACCCCTGGACCGCTGGACGCCACGATGATGTTATGATCAAA 37425
Oy 421 agtaccctcccgagggaatctcaactgacacccctctacacccaattgaag 480
    |||||||
Db 37426 AGGCAACCCCTCTCAGGAATCTCAGCTGACACACTCTACTAGCCCAATTCAGCAG 37485
Oy 481 aagcagtagagcagctgtcagcacaacctcccaacagtaacttggtcttctcgttga 540
    |||||||
Db 37486 AAGCAGTTAGAGCGGTGTGGCGCCACACTCCCAAGACTTAGGTTTCTCTTGGA 37545
Oy 541 gggtagcagtagaagacagagtagctgattctcctagcgtgactaagaatccnaagcct 600
    |||||||
Db 37546 TGGGGGAGCTGAGACAGAGGATAGCTGATTTCTTAGCGTACATAAGATCCCTTAACCT 37605
Oy 601 aacgggaaagtgacgcacatccatcttaaacatggggctgacactgaactacacccg 660
    |||||||
Db 37606 AGCTGGGAAGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTCACACCTG 37665
Oy 661 aacacagagagcagctcaataatgctcaatcagcacaacagaggtaagacatagcc 720
    |||||||
Db 37666 ACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAGAAATAGCC 37725
Oy 721 aatcatctatgctctgagagacagcgaggaaagacagagatlggataataactcaggca 780
    |||||||
Db 37726 AATCATCTAATGCTCTGAGACAGACAGAGGAGCAATGATCGGATTAACCCAAAGTC 37785
Oy 781 ctcaagcagagaagaagaacccctctggggccctccatctgtaaggagctcgtttc 840
    |||||||
Db 37786 TTCCAGCGCGCAACGGCAACCCCTTTGGGGCCCTCCCTTGTATGGGAGCTCTGTTTT 37845
Oy 841 cactctatctcaactctatlaaataatgcaactgacactctcctggttccgttcttcttga 900
    |||||||
Db 37846 CATGCTATTTCACCTAATTAATCTTGCAACTGCACTCTTGCTCATGTTTCTTAACGG 37905
Oy 901 ctcaagctgagctcttctgctgcccacacacacacacacacacacacacacacacac 960
    |||||||
Db 37906 CTTAGGTGACCTTTCCTCCCTCCACACACACCTGCTTTGGCCGACCGAGACCCGCC 37965
Oy 961 gctgaactccacccctctgagatccacagagagtgctcactgctgctccgacacagaggt 1020
    |||||||
Db 37966 GCTGACCTCCATCCCTCTGATCATGCAAGGTGTCCGCTGCTGCTGATCAGCAGAGCC 38025
Oy 1021 acccattgcccacccgacagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
    |||||||
Db 38026 ACCCATTTGCCGCTCCCAATCGCGGCTAAAGGCTTGCATTTGTCCTGATGCGTAAAGTGC 38085
Oy 1081 tgggttgcctctaaagaacacagacactggttcaactggttccatggttctctcctcagac 1140.
    |||||||
Db 38086 TGGGTTCATCTTAATTAAGCTGAACACTAGTCACTGGGGTTCATGATGTTCTTCTGTGAC 38145
Oy 1141 ccaagcgtcttaataagactaatacactcacacgcatgagcccaagatccatctcctgta 1200
    |||||||
Db 38146 CCACAGCTTCTAATAGAGCTAATACTACACCGCATGCGCCCAAGGTTCCATCTT-GAA 38204
Oy 1201 tctgtgagcccaagaaccccaagtcagagaangtgaagcttgcacacatttgaggaa 1257
    |||||||
Db 38205 TCCATTAAGCCCAAGAACCCCAAGTCAGAGAAACAGAGGCTTGCCACCATCTTGGGAG 38261

```

```

RESULT 4
LOCUS HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 56093)

```

AUTHORS Pauley, A.
 TITLE The sequence of H. sapiens BAC clone RG083M05
 JOURNAL Unpublished (1996)
 REFERENCE 2 (bases 1 to 56093)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (13-NOV-1996)
 COMMENT Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63108, USA
 e-mail: saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:
 This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
 VECTOR: pBelo
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H_RG083M05; actual end is at 56093 of H_RG083M05

This clone contains STS SWS1725.

FEATURES

source
 location/Qualifiers
 1..56093
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q21-7q22"
 /clone="H_RG083M05"
 /clone_1lb="CITB-978SK-B"
 complement(838..1131)
 /rpt_family="Alu"
 <1360..16971
 /gene="WUGSC:H_RG083M05.1"
 join(<1360..1503,4181..4370,4587..4774,6422..6556,9483..9547,11631..11773,11864..12021,13131..13296,14885..14988,16349..16546,16837..16971)
 /gene="WUGSC:H_RG083M05.1"
 /note="ATPase; strong similarity to peroxisome biosynthesis protein PAB1 (P1D:q1172019); coded for by human cDNA C04279 (NID:q1467530)."
 /codon_start=1
 /protein_id="AAB46346.1"
 /db_xref="GI:1669371"
 /translation="KRLENIQKLEVAESEAVMOPSVLLDDDLIACLPAVPEHEH
 SPDAEORCELTIVNINKLDCIDNFTDLDLQNAVAKETGFAVDTVLIVRAVPEHEH
 SROSISREKLVLTLDLQFQKALDGLFAPSLRSVNLKRPDLGDKTGLAEHQIILMD
 TLOLPARKVCLKEKPELPAFLPRLPRTGILLYGPGTGTLLAGVIARESHNFISV
 KGPBLSKYIGASEQAVRDLFIETRAQAKPCILFEDEFESIAPRGHDNTGVDRVVO
 LITOLDVSGICGVVLAATSRPDLDPALRPGRDKVCVCPPOVATVSTESKTO
 OMLHSVLVRELINLVSDSLPRLADVDLQHAASVYDFSTGADIKALLVNAOLEAHG
 MLKSKEILPDESKTNMYRLTFGSSTSELGNGTSSDLSQCLSPSSMTDLPSPV
 GKQDLFSQPPVLRTASQEQDELTDQROQLRADLSIKGRVRSQSGEEDSMNQPI
 KRLIASOSHLMVALGHTRPISSEDDMKFAEL"
 complement(4948..5130)
 /rpt_family="Alu"
 complement(6581..7133)
 repeat_region
 repeat_region
 repeat_region

repeat_region	/rpl_family="L1"	complement(7767. .8037)
repeat_region	/rpl_family="ALU"	complement(8186. .8472)
misc_feature	/rpl_family="ALU"	8473. .8625
misc_feature	/gene="WUGSC:H:RG083M05.1"	/note="match to human 3' EST H75782 (NID:g1049794), bases 287-444"
misc_feature	/gene="WUGSC:H:RG083M05.1"	8841. .9161
misc_feature	/note="match to human 5' EST H75921 (NID:g1050050), bases 21-348"	9481. .9547
repeat_region	/gene="WUGSC:H:RG083M05.1"	/note="match to human 5' EST N22627 (NID:g1130501), bases 276-343"
misc_feature	/rpl_family="ALU"	complement(12612. .12907)
misc_feature	/rpl_family="ALU"	13670. .13793
repeat_region	/gene="WUGSC:H:RG083M05.1"	/note="match to human 5' EST H41382 (NID:g17434), bases 143-266"
repeat_region	/rpl_family="ALU"	13794. .13877
misc_feature	/gene="WUGSC:H:RG083M05.1"	13878. .13906
repeat_region	/note="match to human 5' EST H41382 (NID:g917434), bases 30-58"	13907. .14104
repeat_region	/rpl_family="ALU"	complement(14110. .14137)
repeat_region	/rpl_family="L1"	complement(15618. .15907)
repeat_region	/rpl_family="ALU"	17227. .17522
misc_feature	/rpl_family="ALU"	18667. .19235
repeat_region	/note="match to human fetal brain 5' EST D61494 (NID:g9704009), bases 1-255, and to human 3' EST R07476 (NID:g159399)"	19550. .19670
repeat_region	/rpl_family="ALU"	21507. .37303
misc_feature	/note="similarity to various SS-RNA virus polypeptides; pseudogene; region of matches and close matches to multiple human ESTs, see R68740 (NID:g842257)"	37316. .37489
exon	/note="Grai1 prediction, score = 80"	
repeat_region	/evidence=not_experimental	complement(38938. .39224)
misc_feature	/note="match to multiple human ESTs, see N30113 (NID:g1148633)"	39225. .39707
repeat_region	/rpl_family="ALU"	complement(40247. .40538)
repeat_region	/rpl_family="ALU"	complement(40632. .40924)
repeat_region	/rpl_family="ALU"	complement(42283. .42891)
repeat_region	/rpl_family="ALU"	complement(45474. .45613)
misc_feature	/rpl_family="ALU"	complement(45614. .45737)
misc_feature	/note="match to human 3' EST H48898 (NID:g988738), bases 129-333"	
misc_feature	/note="match to multiple human ESTs, see N81064 (NID:g1243765), H48897 (NID:g988737), and M78831 (NID:g273146)"	complement(47027. .47318)
repeat_region		

Query Match	Best Local Similarity	Score	DB	Length
Matches 1133; Conservative	90.1%;	1035.2;	W71912 (NID:g273505)	56093;
		Pred. No. 6.1e-304;		
		Mismatches 111; Indels 13; Gaps 2;		
1	tcanaatcgaagagcttgaacttctgaacgcgcaaaagaggaggaaactgtttatttt	60		
Db 37018	TCAAAATCGAAGAGCTTTAGACTTGTGTAACCGCTGAAAGAGGGGAGACCTGTTATTTT	37077		
Qy 61	aggggagaagatgcgtttagtattcattcaatcaactggaacatctactgagaaagttaaaga	120		
Db 37078	AGGGGAGAAGATGCTGTTATTATTGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA	37137		
Qy 121	aatttgagatgataataatgttaagcagaggaacttcaaaacactgtcacccctggggcct	180		
Db 37138	AATTGAGAGTCAAAACAACTGAGAGCAAGAGAGCTTCCAAACACTGAGACCTGGGGCCT	37197		
Qy 181	cttcagcaaatgagatgccccttggaactcccccctttagaagactctagagcctataatatt	240		
Db 37198	CCTCAGCCAAATGAGATGCCCTTGATTTCTCCCTTCTTGAAGACCTTGAAGCAAGCTATTAATAT	37257		
Qy 241	tttaactcctctcttggaacctgtactctcaacttccctgttgaagtgtgtctctccagaat	300		
Db 37258	GCTACTCCTCTTTGGACCTGTATGCTTTAACTCCTGTTAATCTTGTCTCTCCAGAAAT	37317		
Qy 301	tgaagctgtaaagctacaaatagttctcttcaaatgaaacccagatgagtcctcagactaa	360		
Db 37318	CGAAGCTGTAAACATA-----CAATGTGAGCCCAAGATGACATCCCAAGCACTAA	37365		

```

Oy 361 aatcaccgtagaacccctgagccgctgctagactatgctctgtaattgaacattga 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37366 GATCTACCGCAGACCCCGGACCGGCGCTGATAGCCACAGACTGTATTAATGACATCA 37425
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 421 agtcacccctcccgaggaatctcaactgacaaacctactaacctcaattcaagtga 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37426 AGGCACCCCTCTCTGAGGAAATCTCAGCTGCACAACTTACTACGCCCAATTCAGCAGG 37485
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 481 aagcagcttagagcaatctgtagcccaacctccccaacagtaactggtctctctgttga 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37486 AAGCAGTTAGAGCGGTGCTCGGCCCAACCTCCCAACAGCAGCTAGTTTCTCTGTTGAGA 37545
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 541 gggtagactgagagacaggaactgactgtatctctaggctgaactaagaaacccaagcct 600
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37546 TGGGGGACTTGAGAGACAGAGACTGAGTTCCTAGGCTGACTAAGATATCCCTTAAGCT 37605
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 601 auctggagagtgaccgcatcatctttaacatgaggcttgcaacttaagctcaacaccg 660
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37606 AGCTGGAAAGTGACCACTTCAACCTTAAACAGGGGCTTGCACTTAGCTCACACCTG 37665
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 661 accaatcagagagctcacttaaaatgctaatacagcaaaaagagagagtaaaagcaataagcc 720
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37666 ACCAATCAGAGAGCTCAGCTAAATATGCTAATTAAGCAAGAGAGTAAAGAAATAGCC 37725
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 721 aatcattcatgtcctgagagacagcgaggaggaaggaagattggaataataactagga 780
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37726 AATCATCTATTGCTCTGAGAGCAGACAGAGAGGAAATCGGATTTAAACCCAACTC 37785
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 781 ttcaagcagacagacagacacccctctgtgctccctccatgtatgtggagctcgtttt 840
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37786 TTGCAAGCCGCGCAAGCGCAACCCCTTTGGGTCCTCCCTTTGATGGAGCTCGTTT 37845
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 841 caacttatttcaactatataatcaatgcaactgacactctctgtgcgtgttttattg 900
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37846 CATGCTATTCTACTCTATTAAATCTTGCAACTGCACTCTTGAGTCCATGTTTCTTAAGG 37905
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 901 ctcaagcgtgagcttctgtgctgcatcccaactgctgttggcaaccgtcaagaccgct 960
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37906 CTGAGCTGAGCTTTCGCTCGCATCCACACTGCTGTTGGCCGCGCACGACCCGCC 37965
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 961 gctgactcaccatccctctgtagatccagagagtgctcaactgtctctgatacgaagagt 1020
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 37966 GCTGACTCCCATCCCTCTCGATCATGCAAGGGGTGCTCGCTGCTCTATCAGGAGGC 38025
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1021 accaatgcaactcccgatcaggtcagaaggtctgcaattgtctctgcaatgcaatgctc 1080
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38026 ACCCATTCGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTTGCCATGGCTAAGTCCC 38085
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1081 tgggttctgctaataagaaactggaactggtgcaactggttccatggtctcttccatgac 1140
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38086 TGGGTTCAATCCTAATTAAGCTGAACTGACTGGGTTCATAGTGTCTCTCTGTGAC 38145
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1141 ccaaggtctctaataagaaactaataacactcacgcgaatggcccaagaattccatctctgtga 1200
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38146 CCACAGCTTTTAATAGACTATTAACACTCACCGCATGGCCCAAGTTCCATTTCCTT-GAA 38204
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1201 tctgtgagagccaagaaccacaggtcagaagaangtvgaggttgcacaaatttgggaag 1257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 38205 TCCATTAAGGCCAAGAAACCCAGGTCAGAGAACACAGAGGCTTGCCACCATCTTGGGAG 38261
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 149194)
TITLE Waterston,R.H.
JOURNAL Direct Submission
SUBMITTED (15-MAY-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 149194)
TITLE Waterston,R.
JOURNAL Direct Submission
SUBMITTED (16-NOV-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 149194)
TITLE Waterston,R.H.
JOURNAL Direct Submission
SUBMITTED (03-JUN-2002) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 149194)
TITLE Waterston,R.H.
JOURNAL Direct Submission
SUBMITTED (06-FEB-2002) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Nov 16, 2000 this sequence version replaced g1:4835815.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H.RG010605
-----

FEATURES
Source 1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTB-10G5"

BASE COUNT 45439 a 29540 c 31229 g 42986 t

ORIGIN
Query Match 77.9%; Score 1035.2; DB 9; Length 149194;
Best Local Similarity 90.1%; Pred. No. 6.2e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

Oy 1 tcaaatcgaagagcttagaactgctcaaccgccaagaagaggaggaactgttaatttc 60
Db 84986 TCAAAATCAAGAGAGCTTATGTTAATCAATCCGGAACTGCTCACTGAGAAATTAAGA 84927
Oy 61 aggggaagaatgctgtatgttaataatcaatctggaatcttactgagaagattaaaga 120
Db 84926 AGGGGAAGAATGCTGTATTAATGTTAATCAATCCGGAACTGCTCACTGAGAAATTAAGA 84867
Oy 121 aattgagatgaatataatgtagacagagagactcaaaacactgcaacctggggcct 180
Db 84866 AATTGGAATGGAATGACACTAGACAGAGAGAGAGCTTGGAACACTGAGACCTGGGGCT 84807
Oy 181 cctcagcaatgtagtccctggaactccctctttagagaccttagagcttagagctataat 240
Db 84806 CCTCAGCAATGAGATGCGCTGATTTCTCCCTCTTAGAGACCTTAGAGAGCTAATAATTT 84747
Oy 241 ttactctcttggaccctgtacttcaacttccctgtgtgaagttgtcttccagaat 300

```

|||||
Db 84746 GCTACTCCCTTTGGACCTGTATCTTTAACTCCTTTTAATTGTCTCTCCAGAA 84687
QY 301 tgaagctgaaagctacaagaatagttcttaaatgaaaccccgagatccatgactaa 360
Db 84686 CGAAGCTGTAAACCTA-----CAATTTGGAGCCCAAGATGCACTCCCAAGACTAA 84639
QY 361 aatctacccttgagacccttgagccgctgtgtagactgtctgtagtgaatgacatga 420
Db 84638 GATCTAACCGCAACCCCTTGACCGGCTCTAGCCACAGATGTGATGATTAAAGACATCA 84579
QY 421 agtaccctcccgaggaaatctcaatgacaaacccctactactacatccatccagtag 480
Db 84578 AGGCACCCCTCTGAGGAATCTCAAGCTGCACAACTCTACTAGCCCAATTCACAGAG 84519
QY 481 aagcagttgagcagttgtcagccaaacctcccaaaagtaacttggttttctgttgaga 540
Db 84518 AAGCAGTTAGAGCGGTGCGGCGCAACCTCCCAACAGCACTTAGGCTTTCTGTGGAGA 84459
QY 541 ggttgagctgagagacagagactgtagttccctagagctgagtaagaaatccnaagcct 600
Db 84458 TGGGGGACTGAGAGACAGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 84399
QY 601 anctggaaagtgacgcgacatctttaaacaatggggcttgcaacttagtcaacacccg 660
Db 84398 ACCTGGAGAGGTGACCACTCCACCTTTAAACAGGGGCTTGCAACTTAGCTACACCTG 84339
QY 661 accaatcagagagctcaccataaagttaacagcaagagagagagagagagagagagag 720
Db 84338 ACCAATCAAGAGCTCACTAAATGCTAATAGCCAAAGAGAGAGAGAGAGAGAGAGAG 84279
QY 721 aatctactatgctgtagagcagcagagagagagagagagagagagagagagagagag 780
Db 84278 AATCATCTATTGCTGTAGAGGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 84219
QY 781 ttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
Db 84218 TTCCAGCCCGCAACGCAACCCCTTTGGATGCTCCCTCTTGTATGAGAGCTCTGTTT 84159
QY 841 cactctatctactatataatcaatcaatcaatcaatcaatcaatcaatcaatcaatca 900
Db 84158 CATGCTATTTCATCTATTAAATCTTGCAACTGCACTCTTCTGCTGCTCAATGTTTAC 84099
QY 901 ctcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960
Db 84098 CTTGAGAGCTTTCGCTGCGCATCCACACAGCTGTTTGCGCGCAACGCAAGACCGCC 84039
QY 961 gctgagcttcaatcccttgagcagcagcagcagcagcagcagcagcagcagcagcag 1020
Db 84038 GCTGACTCCCATCCCTGATGATCAGAGGCTGCTCGCTGCTGCTGATTCAGCAGCAGC 83979
QY 1021 accaatgccaatcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1080
Db 83978 ACCCAATGCGCGCTCCCAATCGGGCTTAAAGGCTTGCCATGTTCTCAATGAGCTAAG 83919
QY 1081 tgggttctgcttaataagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg 1140
Db 83918 TGGGTTCATCTCAATGAGCTGAACCTAGTCACTGGGTTCATGTTCTCTCTTGAC 83859
QY 1141 ccaagcctctaaatagactataaactcaacgacgacgacgacgacgacgacgacgacg 1200
Db 83858 CCACAGCTCTTAATAGATTAAGCTTAAGCACTACCGCATGGCCCAAGGTTCCATTCCTT 83800
QY 1201 tctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1257
Db 83799 TCCATTAAGGCCAAGACCCCAAGTCAAGAAACAGAGGCTTGCCACCACTTCTGGAG 83743

RESULT 6
AX007980 10499 bp DNA linear PAT 06-SNP-2000
LOCUS AX007980
DEFINITION Sequence 3 from Patent WO967395.
ACCESSION AX007980

VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 10499)
TITLE Perin, J.P., Rieger, F. and Alliel, P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
location/Qualifiers
FEATURES
source 1..10499
BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN
Query Match 77.0%; Score 1023.2; DB 6; Length 10499;
Best Local Similarity 90.1%; Pred. No. 2.7e-300;
Matches 1132; Conservative 0; Mismatches 111; Indels 14; Gaps 3;
QY 1 tcaaaatgaaagagcttagacttgctaacgcgcaaaagaggaggagagcgtttatttt 60
Db 9018 TCAAAATCGAAGAGCTTTGACTGTGAACCTGAAAGAGGGGAACCTGTTATTTT 9077
QY 61 aggggaagatgctgttagtatgtaatacatctggaatcattactgagaaagttaaaga 120
Db 9078 AGGGGAAGATGCTGTATTATGTTAATCAATCCGAATCGACACAGAAAGTTAAAGA 9137
QY 121 aattgagatcgaataatgtagagcagagagagccttcaaaacacgcacccctgggct 180
Db 9138 AATTGAGATCGAATTAACAAGTAGAGACAGAGAGCTTCCAAACACGACCTGGGCGCT 9197
QY 181 cctcaagcaatgtagctccttgagcctcctctcttagagcctctgagcagcagcagcagc 240
Db 9198 CCGACACCAATGATGATCCCTGATGTTCTCCCTTTAGAGACCTCTAGACGCTAATATT 9257
QY 241 tttaactcctcttgagccctgtatcttcaactcctcttgtaagttgtctctcagaat 300
Db 9258 GCTACTCCCTTTGGACCCCTGATATCTTAACTCTTGTAACTTGTCTTCCAGAAAT 9317
QY 301 tgaagctgaaagctacaagaatagttcttaaatggaaccccgagtcagttccatgactaa 360
Db 9318 CGAAGCTGTAAACCTA-----CAATGAGAGCCCAAGATGCACTCCCAAGACTAA 9365
QY 361 aatctacccttgagcccttgagccgagcgtgtgtagactgtctgtagttaaagacatga 420
Db 9366 GATCTACCGCAACAGCCCTGAGCGGCTGCTAGCCACAGATCTGATGATTAAAGACATCA 9425
QY 421 agtaccctcccgaggaaatctcaatgcaaaacccctactactacatccaatccagtagg 480
Db 9426 AGGCACCCCTCTGAGGAATCTCAAGCTGCACAACTCTACTACGCCCCAATTCACAGG 9485
QY 481 aagcagttgagcagttgtcagcgaacccctcccaaaagtaacttggttttctgttgaga 540
Db 9486 AAGCACTTAGAGC-GGTCTCGGCCAACCTCCCAAGAGCACTTAGGTTTCTGTTGAGA 9544
QY 541 ggttgagcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 600
Db 9545 TGGGGGACTGAGAGACAGACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCT 9604
QY 601 anctggaaagtgacgcgacatccatcttaaaacatggggcttgcaacttagtcaacccg 660
Db 9605 AGCTGGAGAGGTGACCACTTAAACAGGGGCTTGCAACTTAGCTACACCTG 9664
QY 661 accaatcagagagctcacaataatgtaatacagcgaagagagagagagagagagagagag 720
Db 9665 ACCAATCAGAGAGCTCACTAAATGCTAATTTAGGCNAAGCAGAGAGTAAGAAATAGCC 9724

Query Match 75.7%: Score 1006.6: DB 2: Length 188919;
 Best Local Similarity 86.7%: Pred. No. 3.2e-295;
 Matches 1160: Conservative 0; Mismatches 157; Indels 21; Gaps 4;

QY 1 tcaaaatcgaagagctttagacttgcgaacccgcaaaagaaggaggggaactgttatttt 60
 Db 22917 TCAAAATCGAAGATTTTAACTGGTAAACACCAAAAGAGGGGAAACCTGTTATTTT 22858
 QY 61 agggagaagaatgctgttagtagtaatcaatcgtgaatcaatcgtgaagaagttaaaga 120
 Db 22857 AGGGGAACATGCTGTATATATGTTAATCAATCTGATCATCAGGAGGAACTTAAAGT 22798
 QY 121 aatcttagatcgaataaagttagagagagagagagagagagagagagagagagagag 180
 Db 22797 AATTCAGATCGAATACATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 22738
 QY 181 cctcagccaatgagatgag 240
 Db 22737 CTTCAGGCAATGAAACATGAGATGTCCTCCCTCTTAGAGACCTAGCAGCTATATATAT 22678
 QY 241 ttaactcctcttggag 300
 Db 22677 GTCACTCTCTTGGAGACATGTAATTAACCTCTGTTAAGTTGTTCTTCCAGAAAT 22618
 QY 301 tgaagctgtaaaagctaaatagttcttcaaatggaacccagagatgagatcagagataa 360
 Db 22617 TGAAGCTGTAAGCTTAACAAATGTTCTTCAAAATGAGCTCAAGATGAGTCATAGCTAA 22558
 QY 361 aatcaccgtgag 420
 Db 22557 AATCTGGCGGGAGACCTTGAGACAGCTGCTAGCCCATGTCGATGTTAATGACATCA 22498
 QY 421 agtcaacccctccggaggaaatctcaatgcaaaacccctactacatccaatcagtagg 480
 Db 22497 GGGGACCCCTCTGAGAAATCTCAACTGCAACACCCCTACTATGCCAGTTCCAGAGG 22438
 QY 481 aagcagtttag 540
 Db 22437 AAGCAGTTAAGTGTGATGTCAGGCAACCTCCCAACAGCACTTGCTTTCCTTACGA 22378
 QY 541 ggggtgagctgag 600
 Db 22377 GGGGGAGTCTGAGACACAGATGTTAGTGTGATGTTCTAGCGCAGCTAAAGATCCCT 22318
 QY 601 anctggagaaggtgag 660
 Db 22317 AGCTGGAGAGGTATCATCATCCACTTTAAACATGGGGCTTGCAACTTACTCACACCG 22258
 QY 661 accaatgag 712
 Db 22257 ACCAATAGAGTATGAAGAAGCTCACTAAAGCTATTGGCAAAAACAGAGAGGTTAAG 22198
 QY 713 caataagcaatcatctatctgcttggagagagagagagagagagagagagagagagagagag 772
 Db 22197 AAATAGGCATATCGCTATGCTGAGAGACAGATGGAGGAGCAATGATCAGATATATA 22138
 QY 773 ctaagagatcgaag 831
 Db 22137 CCCAGGATTCGAGCGGCAATGCTTAAACCGCTTTGGGTCCTCCCTTGTATGGGAG 22078
 QY 832 ctctgtttcaactctatctatctatcaatcaatcaatcaatcaatcaatcaatcaatcaat 891
 Db 22077 CTCTGT-----TTTCACTATATTAATCTTGCACATGCAATCTCTT--CTGGTCC 22030
 QY 892 ttattatgctcaagctgag 951
 Db 22029 GTGTATAGGCTCTAGTGAAGCTTTCGCTGCGCATCCACACTGTGTTGCGCGTGTGC 21970
 QY 952 agaccgctgctgag 1011
 Db 21969 AGACCCGCCACTGACTTCCATCTCTGGGTCGGGAGGAGGTCCGCTGCTCTCTGATC 21910

QY 1012 cagcagagtagaccatctgaccatcccatcagcagagagagagagagagagagagagagag 1071
 Db 21909 CAGCCAGGACCCATTCCTCTCTCGATCGGGCTTAAGGCTTGCCATGTTCTCTGACGG 21850
 QY 1072 ctatgctgctgggtctgttctcctaataagagagagagagagagagagagagagagagagagag 1131
 Db 21849 CGAAGTCCCTGGGCTTCATCTAATGAGCTGAAATACAGTACATGAGTGGTTCACAGGTTCTC 21790
 QY 1132 ttccatgag 1191
 Db 21789 TTCCATGACCCATGCTCTTAAATAGACCTTAACATCACCACCGATGCGCCCAAAATTCAT 21730
 QY 1192 tccctgtatctgtgaag 1251
 Db 21729 TCCTTGAATCTGTGAGGCGCAAGACCCAGTGTCAGAGAACAGAGAGCTTGCCACATCT 21670
 QY 1252 gggagagtgagcag 1311
 Db 21669 TGAAGTGGCCACACACATCTTGAAGTGCGCTGACCATCTTGAGAGCTCTGGGAGC 21610
 QY 1312 aagagatcccccagataa 1329
 Db 21609 AAGGAACTGCCAGTAGCA 21592

RESULT 8
 AC092708/c 161571 bp DNA linear HTG 20-JUL-2001
 LOCUS
 DEFINITION Homo sapiens chromosome 18 clone RP11-141A18 map 18, WORKING DRAFT
 SEQUENCE, 3 unordered pieces.
 AC092708
 AC092708.1 GI:14971454
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161571)
 Homo sapiens chromosome 18, clone RP11-141A18
 2 (bases 1 to 161571)
 Unpublished

REFERENCE
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
 Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Bouhagalter,B.,
 Brown,A., Camarato,J., Campoliano,A., Chang,J., Chazaro,B.,
 Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
 Jones,C., Kamet,A., Karatas,A., Kells,C., Labrecque,K.,
 Lamazares,R., Landers,T., Lehocaky,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
 Menius,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Sounguez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,
 Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
 Direct Submission
 JOURNAL
 Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12749
Center clone name: 141A.18
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16083 bases at least Q40
Consensus quality: 161248 bases at least Q30
Consensus quality: 161352 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 161371; sum-of-ctrls
Quality coverage: 9.8 in Q20 bases; sum-of-ctrls
Quality coverage: 9.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 32930: contig of 32930 bp in length
* 32931 33030: gap of 100 bp
* 33031 114301: contig of 81271 bp in length
* 114302 114401: gap of 100 bp
* 114402 161571: contig of 47170 bp in length.
Location/Qualifiers
1. 161571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-141A18"
/clone_1b="RPC1-11 Human Male BAC"
1. 32930
/note="assembly-fragment"
clone_end:SP6
vector_side:left
33031..114301
/note="assembly-fragment"
114402..161571
/note="assembly-fragment"
clone_end:T7
vector_side:right"
E COUNT 45518 a 33859 c 34317 g 47677 t 200 others
IGIN
Query Match 74.5% Score 989.6; DB 2; Length 161571;
Best Local Similarity 88.6%; Pred. No. 4.9e-290;
Matches 1146; Conservative 0; Mismatches 117; Indels 30; Gaps 6;
QY 1 tcaaaatcgaagagcttgaacttgaaccgcaaaagagggagacgttattt 60
DB 23281 TCATAATTGAAGAGCTTTAGACTTGTCTGCTAACCAACCAAGAGGAGACCTTTATTATTTT 23222
QY 61 aggggaagaatgctgttagttagttaaataatcgtgaatcatctcctgagaagttaaaga 120
DB 23221 AGGGGAAGAAAGCTGTATTATTGTTAAATCAATCCGGAATCATCCGAGAAAGTT---A 23166
QY 121 aattgagaatgaataatgaatgaagagagacctaacaacactgagacctggagct 180
DB 23165 AATTCAAGATCAAAATACATATAGACGAGAGAGCTTCAAAACACTGGACCTGGGCT 23106
QY 181 cctacgaacgaatgctccttgaactcctccttctgaagacactgaagactataat 240
DB 23105 CCTCAGCCAAATGAGATGCCCTGATTCCTCCTTTAGAGACTCTAGAGAGCTATATATT 23046
QY 241 tttaactcctcttgagaccgtatcttaactcctctgttaagttgtctctccagaat 300

|||||
DB 23045 GCTACTCCTTTTGGACCTGTATCTTAACTCCTTTGTTAAGTTGTCTTCAGAAAT 22986
QY 301 tgaagctgaagaagcttaacaaatagttcttcaaaatgaacccagaatgagctcagactaa 360
DB 22985 TGAAGGTGAAGTACAAATCAATTTCTCANATGAGGCCAGATGCAATGCAATGACTTAA 22926
QY 361 aatcaccgttgagacccttgagaccgctcgtcagactaactcgtatgttaatgaatga 420
DB 22925 AATGTACCGTGGACCCCTGGACACAGCCGTGTAAGCCATGCTCAATGTTAATGACATCAA 22866
QY 421 agtaccctcctcgaaggaatctcaactgcaaacccctactacatccaatgaatga 480
DB 22865 AGGACCCCTCCCAAGAAATCTCAACGCAACACCCCTACTTACACCCCAATTCAGCAGG 22806
QY 481 aagcagttagaagagctgttgagccaacctcccaacagtaactgtggttctctgttga 540
DB 22805 AAGCAGTTAGACAGCTTTGTACCCCAACCTCCCAACACACTTTGGTTTCTCTTGAGA 22746
QY 541 ggtgagctgagagacagagactagctgtgattccttaggtcgtactaagaatccnaagct 600
DB 22745 GGGGGAGACTGAGACAGAGACTAGCTGATTTCTTAGGCCAATTAAGATCCTTAAGCCT 22686
QY 601 aactggaaggtgacgcgaatccatcttaacaatggtgctgcaactagctcaaccg 660
DB 22685 AGCTGGGAAGGTGACCGCATCCACTTTAAACATGGGGCTTGCAACTTAAGCTCACACCGG 22626
QY 661 ac-----caacagagagctcactaaatgcttaatcagagcaaaaacagagtaag 712
DB 22625 CCAACAGGTAAATTAAGAGAGCTCACTAAATGCTAATAGGCAAAAACAGAGTAA 22566
QY 713 caatagcaatcatalctatctgctgagagcagcgagagagcaagatctgagataaa 772
DB 22565 AAATAGCCAAATCATCTTTGCGCTGAGACACAGGGGGGAGCAATGATCAGATATATA 22506
QY 773 ctcaagcaatcaagcagcaacaaaccccttgggttccctcccatctgatatggagc 832
DB 22505 CCCAGGCAATTCACCAACCAATGCTACCTTTGGTCCCTCTTCTTATGGGAGC 22446
QY 833 tctgttcaactcactcactcactaaatcaatgcaact--ggaactctctgttgcgtg 890
DB 22445 TCTGT-----TTTCGCTATTAATTAATCTTGCAACTGCGCTCTTGTGGTGCATG 22396
QY 891 tttttat-----gctcaagctgagcttctgttcgcacatccacactgctgttgcac 945
DB 22395 TTTGCTATGCTCGGCTCTAGCTGAGCTTTGTTACATGTCACACACGCTGTTTGGCTGC 22336
QY 946 cgtcaagaccgctgctgacttccatcccttctgatacgaagatgttcaactgtgtc 1005
DB 22335 CGCGCAGACCTGACACTGACTTCCATCCCTTCGATTCGAGATTCGAGGATCTCTGTGCCC 22276
QY 1006 ctgatacgaaggtlaacactatgcaatcccatcagactaaagcttgcatctgttcc 1065
DB 22275 CTGATCCAGCAAGGACCACTGCTGCTCGTCGATCAGGCTAAAGCTTGCCATGTTCT 22216
QY 1066 gcatgctgaagtgctggttcttctaaatagaactgaacatgctgacgtgttccatg 1125
DB 22215 GCATGGCTTAAGTGGCCAGAGTCACTCAATGAGCTGAACACTGATGCTAGTGGTTC--CA 22157
QY 1126 gtctcttccatgaacccagctctctaaatagaactataaacactaacgcatgagccaa 1185
DB 22156 GTTCTCTTCCTGACCCCTGCTCTTAATTAAGCTATTAACACTCACTCAAGGCCCAAGA 22097
QY 1186 ttcaatccttgatctgtaggccaagaccccaagctcagagaagatgagagcttgcca 1245
DB 22096 TTCCATTTCTTGGAATCATATAGGCAAGAACCCAGTCAAGAACACAGAGCTTGCCA 22037
QY 1246 ccatttgggaagtgccccactgcatcttggtagc 1280
DB 22036 CCATTTTGAAGTGCGCTGCTCATCTTGGAGC 22002

RESULT 9


```
repeat_region /rpt_family="MER103"
complement(19991..20119)
/rpt_family="MIR"
complement(20181..20673)
/rpt_family="MLT1G3"
complement(21649..21939)
/rpt_family="Alusx"
complement(22028..22245)
/rpt_family="L1MB8"
22383..22658
/rpt_family="MLT1A1"
22689..22776
/rpt_family="MLT1A1"
complement(23084..23320)
/rpt_family="L1ME4A"
23541..23586
/rpt_family="(ATG)n"
complement(24127..24226)
/rpt_family="MIR3"
24560..24769
/rpt_family="MIR3"
25483..26173
/rpt_family="MER21B"
complement(26203..26366)
/rpt_family="MIR"
complement(26379..26779)
/rpt_family="MSTB"
27372..27438
/rpt_family="(CAG)n"
28669..29257
/rpt_family="MER4B"
29324..29465
/rpt_family="MIR3"
29489..29841
/rpt_family="THE1B"
complement(29844..30059)
/rpt_family="MER20B"
complement(30458..30834)
/rpt_family="MER65A"
complement(30935..31258)
/rpt_family="Aluub"
complement(32849..33446)
/rpt_family="T1ger3(Golem)"
complement(33448..33750)
/rpt_family="Alusg"
complement(33778..34403)
/rpt_family="PABL_B"
complement(34432..34481)
/rpt_family="PABL_B"
complement(34482..34531)
/rpt_family="T1ger3(Golem)"
complement(34540..34639)
/rpt_family="T1ger3(Golem)"
34740..34928
/rpt_family="MER58A"

Query Match 74.58; Score 989.6; DB 9; Length 166700;
Best Local Similarity 88.6%; Pred. No. 4.9e-290;
Matches 1148; Conservative 0; Mismatches 117; Indels 30; Gaps 6;

Db 1 tcaaaatcgaagccttagactcgtcgaacgcgaaggaaggggaagccttattttt 60
42844 TCANAATTGAAGACCTTATGACTTGTACCAACAAAGAGAGGGAACCTGTTATTTT 42903

QY 61 aggggaagaatcgtctagtatgttaatcaatctgaaatcaatctagaagaatgaaga 120
|||||
Db 42904 AGGGGAANAATGCTTTATTTATGTTAATCATCCGGAATCATCAACCGAANAATT----A 42959

QY 121 aatttgagatcgaatataatctagagcagaggaaccttcaaaacactgcacccctgggacct 180
|||||
Db 42960 AATTCAAGATCAAAATACAAATAGAGCAGAGAGCTTCAAAACACTGAGACCTGGGGCCT 43019

QY 181 cctcagcaatcgaatgccttgactctccctctcttaggaccttagagactataatatt 240
```

```
|||||
Db 43020 CCTCAGCCAAATGATGACCTCGATTCCTCCCTTTAGAGACCTTAGAGCTAATAATAT 43079

QY 241 tttaactctcttggaacctgtatcttcaacttctcttgtaagtgtgtcttccaagat 300
|||||
Db 43080 GCTACTCTCTTGTGAGACCTGTATCTTTAACTCTCTGTTAAGTTGTCTCTCCAGANT 43139

QY 301 tgaagctgtaaaagctacaatagttcttcaaatgaaaccccaatgacatcactaa 360
43140 TGAAGGTATAAGCTACAAATCATTTCTTAATGAGAGCCAGATGAGCTCATGACTAA 43199

QY 361 aatctacccgtggaacctcgtgacgcgctcgttagactatgctctgatttaataatcga 420
Db 43200 AATCTACCGTGTGACCCCTGAGCCAGCCGCTGAGCCCAATGTTAATGACATCAA 43259

QY 421 agtcaacccctcccggaagaatctcaactgcaaaccccttaatacctcaatcgaatga 480
Db 43260 AGGCAACCCCTCCCAAGGAATCTCACTGCAACACCCCTTACACCCCAATTACGACG 43319

QY 481 aagcagttagaagctgtctagccaacctcccaacagtaacttgggtttccctgttga 540
43320 AAGCAGTTAGAGCAGTTGTGTAGCCAACTCCCAACAGCAGCTGTGGTTTCTGTGAGA 43379

QY 541 ggttggaactggaagacaggaactggaatttctctaggtctgaataagaaatcccaagcct 600
Db 43380 GGGGGAGCTGAGAGACAGAGACTGACTGATTCTTAGGCCCAATGAAGATCCTAAGCCT 43439

QY 601 aactgggaagtggaacgacatcactttaaactgggggtcgtgaacttagtcaacccg 660
Db 43440 AGCTGGGAAGGTGACCCGATCACCCTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 43499

QY 661 ac-----caatcagaagctcaactaaatgtctaactgaagcaaaaaacaggaagtaag 712
43500 CCAACAGGTAAATGAAGAGAGCTCACTAAATGCTAATCAGGCAAAACACAGAGTAA 43559

QY 713 caatagccaatcatctatgtctgaagcagaagcgggaagaaagatttgatataaa 772
Db 43560 AAATAGCCAAATCATCTTTCGCTGAGACACAGTGGGGGAGCAATGATCGAATATAA 43619

QY 773 ctacagcatcaagcagcaacaacacccctcttggttccctcccaattgtaagggagc 832
Db 43620 CCCAGGCAATTCAACCCAGCAATGGCTTACCCCTTTGGGTCCCTCCCTTTTATGGGAGC 43679

QY 833 tctgttttactctattcactcattaaatcatgaact--gaactctctgtccgtg 890
Db 43680 TCTGT-----TTTCGCTTATTAATCTTGCAACTCGGCTCTCTGTGTGATG 43729

QY 891 ttctttat-----ggtcgaagctgaagcttgttcgcacatccacatgctgttcgcac 945
Db 43730 TTTGCTATGGCTCGGCTCTTACCTGAGCTTTGTTCACTGTCACACAGCTGTTTGCTGC 43789

QY 946 cgtacagaacccgctgtcgtacatccatcccttctgatacgaagagtgatcactgtc 1005
Db 43790 CGCGGAGACCTGCGACATGACTTCATCCATCCCTTCGGAATTCGAGAGGTCCTGTGCGCC 43849

QY 1006 ctgatcagcagaggtatcccatctgcactcccgatcaggtcctaagagctgtccattgtcc 1065
Db 43850 CTGATTCAGCAAGGACCCCATTTGCTGCTGATGAGCTGAATCTCTTGCATGTTTCT 43909

QY 1066 gcatgagcgaagctgcgtggttctgcttaatagaactgaacactgtgtcactgtgtccatg 1125
Db 43910 GCATGGCTAATGCTCCAGGTTTCATCTTAATTGAGCTGACACTGATCTGTGCTT--CA 43968

QY 1126 gtctctctcattgacccaagctctcctaataagactataaactcaccgactgtgccaa 1185
Db 43969 GTTCTCTTCCCTGAGCCATGCTCTTCTAATAGAGCTATTAACCTCACTGCAATGGCCAA 44028

QY 1186 ttcaatctctgtatctgtgtgaagccaagaaccccaaggtcagaanaangtgccttgcca 1245
Db 44029 TTTCCATTCCTTGTGAATCATGATGAGCCAAAGAACCCAGGTCTAGAGAACGAGGCTTGCCA 44088

QY 1246 ccatttgggaagtggccacactgtccatttggtaagc 1280
|||||
```

Db 44089 CCAATTTGGAGTGGCCTGCTGCCATCTTGGGAGC 44123

RESULT	10
AP001488/c	
LOCUS	AP001488
DEFINITION	Homo sapiens chromosome 18 clone RP11-693N15 map 18q12, WORKING DRAFT SEQUENCE, 12 unordered pieces.
ACCESSION	AF001488
VERSION	AF001488.3
KEYWORDS	GI:1988499
SOURCE	HTG; HTGS PHASE1; HTGS DRAFT.
ORGANISM	Homo sapiens DNA, clone:RP11-693N15. Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 186911)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 186,911 genomic DNA of 18q12
Published Only in Database (2000) In press
2 (bases 1 to 186911)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical

COMMENT On Jul 14, 2000 this sequence version replaced gi:8117346.

Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gscc.riken.go.jp
Project Information
Center project name: HumDraflt8
Center Clone name: RP11-693N15
Summary Statistics

Sequencing	PCR products	100% of reads
Chemistry	Dye-terminator ET-amersham	100% of reads
Assembly program	Phrap	version 0.990329
Consensus quality	182485 bases at least	Q40
Consensus quality	184520 bases at least	Q30
Consensus quality	185300 bases at least	Q20
Insert size	188111	sum-of-configs
Quality coverage	9.41k in Q20 bases	sum-of-configs

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

38569	6616	38468	bp	in	length
91873	contig of	27548	bp	in	length
91874	contig of	25657	bp	in	length
112197	contig of	20123	bp	in	length
112197	contig of	20340	bp	in	length
147367	contig of	14755	bp	in	length
132632	contig of	17155	bp	in	length
174530	contig of	6814	bp	in	length
1647117	contig of	6821	bp	in	length
171631	contig of	4558	bp	in	length
178852	contig of	3146	bp	in	length
183210	contig of	456	bp	in	length
186456	contig of				

* NOTE: THIS is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown
* This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

*	1	38468:	contig of 38468 bp	in length
*	38469	38568:	gap of 100 bp	
*	38569	66116:	contig of 27548 bp	in length
*	66117	66216:	gap of 100 bp	
*	66217	91873:	contig of 25657 bp	in length
*	91874	91973:	gap of 100 bp	
*	91974	112096:	contig of 20123 bp	in length
*	112097	112196:	gap of 100 bp	
*	112197	132536:	contig of 20340 bp	in length
*	132537	132636:	gap of 100 bp	
*	132637	147361:	contig of 14725 bp	in length
*	147362	147461:	gap of 100 bp	
*	147462	164616:	contig of 17155 bp	in length
*	164617	164716:	gap of 100 bp	
*	164717	171530:	contig of 6814 bp	in length
*	171531	171630:	gap of 100 bp	
*	171631	178451:	contig of 6821 bp	in length
*	178452	178551:	gap of 100 bp	
*	178552	183109:	contig of 4558 bp	in length
*	183110	183309:	gap of 100 bp	
*	183310	186355:	contig of 3146 bp	in length
*	186356	186455:	gap of 100 bp	
*	186456	186911:	contig of 456 bp	in length.

FEATURES	Location/Qualifiers
source	1. .186911

```
misc_feature      /organism="Homo sapiens"  
                  /db_xref="taxon:9606"  
                  /chromosome="18"  
                  /map="18q12"  
clone_RP11-693N15  
1..38468  
/note="assembly_fragment"  
38569..66116  
/note="assembly_fragment"  
66217..91873  
/note="assembly_fragment"  
91974..112096  
/note="assembly_fragment"  
112197..132536  
/note="assembly_fragment"  
132637..147361  
/note="assembly_fragment clone_end:T7 vector_side:left"  
147462..164616  
/note="assembly_fragment"  
164717..171530  
/note="assembly_fragment"  
171631..178451  
/note="assembly_fragment"  
178552..183109  
/note="assembly_fragment"  
183210..186355  
/note="assembly_fragment"  
186456..186911
```

BASE COUNT	55081 a	38409 c	38950 g	53371 t	1100 others
ORIGIN					

Query Match	74.58;	Score 989.6;	DB 2;	Length 186911;
Best Local Similarity	88.68;	Pred. No. 4.9e-280;		
Matches 1148;	Conservative	0;	Mismatches 117;	Indels 30; Caps 6

Oy 1 tcaaatcgaagaagcttagactgtctaaccgccaagaaggggaaacctgtttatttc 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25843 TCAAAATTTGAAGAGCTTTAGACTTGCTTAACCAACCACAAAAGAGAGGCAACCGTTATTATTTTT 25784

```

Oy      61 aggggaagatcgtgtcagtaatcaatctggaatcattactgagaagttaaga 120
          |||||
Db 25783 AGCGGAAGCATGCTGTTATTATGTTAATCAATCCGGAATCATCACCGAGAAAGTT---A 25728

```

121 aattcgatcgaatataatgttagagcagagaccttcaaaaacatgcacccctgggct 180

```

Db 25727 AATTCAGATCAAAATACATGAGACAGAGAGCTTCAAAACACTGAGACCTGGGGCT 25668
Qy 181 cctcagcaatgagatgacctgagacctccctcttaagaccttaagacgtatataatc 240
Db 25667 CCTCAGCCAAATGATGAGCTTGATTTCTCCCTCTCTAGGACCTCTAGACGATATATATT 25668
Qy 241 ttactctctcttgagccctglatctcaactctctgttaagttgtctctcttcagaat 300
Db 25607 GCTACTCTCTTTGGACCTGATATCTTAACCTCTTGAAGTTTGTCTCTTCAGAAAT 25548
Qy 301 tgaagcttataagctacaaatagttctcaatagaaaccagatgagtgatcagactaa 360
Db 25547 TGAAGGTATAAGCTACAAATCATCTTCTCAATGAGGCCAGATGACGTCATGACTAA 25488
Qy 361 aatcactgctgagcccccggagccgctcctcagactatgacctgtatgttaatgactga 420
Db 25487 AATCTACCGTGGACCCCTGGACACGCTGCTAGCCCATGCTCCAAATGTTAATGACATCAA 25428
Qy 421 agtcaaccctcccgaggaatctcaactgcaaacccctactacaatctcaatctagtag 480
Db 25427 AGGCAACCCCTCCCAAGAAATCTCAACTGCAACCCCTACTACACCCCAATTGACAGAG 25368
Qy 481 aagcagtttagagcagttgtctgagccaaactccccaagtaacttgggtttctctgtaga 540
Db 25367 AAGCGTTTAGAGCATGTTTCAAGCAACCTCCCAACACACTTGGGTTTCTGTTGAGA 25308
Qy 541 ggttgagctgagagacagactagctgtatctcctagctgactaagaatccnaagct 600
Db 25307 GGGGGGACTGAGAGACAGACTAGCTGATTTTCCAGGCCAACTAAGATCCCTTAAGCT 25248
Qy 601 anctggaaggttgacgcgacatccatttaacaatggggtgtgaaactagctacacccg 660
Db 25247 AGCTGGGAAGGTGACCGCATCCACTTTAAACATGGGGCTTGCACCTTAAGCTCACACGG 25188
Qy 661 ac-----caatcagagagctcaataaagtcaatcagcaaaacagaggttaag 712
Db 25187 CCAACAGGTAATAAGAGAGCTCACTAAATGCTAATCAGGCAAAACAGAGGTAATAA 25128
Qy 713 caatcagcaatcatalctatctgctgagagacagcggaagaaagaaatgtgatatataa 772
Db 25127 AAATAGCCAAATCATCTTTCGGCTGAGACACAGTGGGGGACAAATGATAGATATAA 25068
Qy 773 ctcaaggaatcaagcagcaaacgcaaccccttgggttccctccatctgtagagagc 832
Db 25067 CCGAGGCAATTCAGGACCAATGCTACCTTTGGGTCCCTCTTGTATGGAGC 25008
Qy 832 tctgtttcaactctatctcaactctataatcatgcaact--gcaactcttctgtccgtg 890
Db 25007 TCTGT-----TTTCGCTCTATTAAATCTTGCAACTGCGCTCTTCTGGTGATG 24958
Qy 891 ttttttla-----ggtcaaaagtgaacttctgtctgcacatccaaactgtctgttcac 945
Db 24957 TTTGCTATAGGCTCGGCTCTACTGAGCTTTTGTTCACATGTCACACATGCTGTTTGCATC 24898
Qy 946 cgtcaagaaacccgctgtcgaactccatcccttggatccagcagagvtctcaactgtctc 1005
Db 24897 CGCGCAACACCTCGCAGATTCATCTCCATCTTCGGAATCGGCAAGGTCTCTCTGTCGCC 24838
Qy 1006 ctgatacagagagtaacatctgcaactccgatacagcctaagagctgtgcatcttctc 1065
Db 24837 CTGATCCAGCAAGGACCACTATGCTGCTCTGATCAAGGCTAAAGCTTGGCCATGTTTCT 24778
Qy 1066 gcaatgctaaagtgcctgggttctgtcttaataagaaactgtaactgtctgttctcaatg 1125
Db 24777 GCATGGCTAAGGCCCAAGTTCATCTTAATGAGCTGAACACTAGTCACTGGGTT--CACA 24719
Qy 1126 gttctctctcatgaccaaagcctcttaataagactataaactcaacgcagatgcccaga 1185
Db 24718 GTTCTCTTCCGTGACCCATGCTCTTAATAGAGCTATTAACACTCACTCAATGGCCCAAGA 24659
Qy 1186 ttcacatctctgtatctgtgagggcagaagcccaaggttcagagaangtgaggtctgca 1245

```

```

Db 24658 TTCCATTCTCTTGGAATCCATGAGCCCAAGAACCCAGGCTGAGAGAACAGAGGCTTGCCA 24599
Qy 1246 ccaattgggaagtggcccaactgacatcttggttagc 1280
Db 24598 CCATTTTGGAGTGGCCGTGCTGCATCTTGGGAGC 24564

RESULT 11
AP001599
LOCUS 118241 bp DNA linear PRI 29-APR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21, clone: KB126A3, App-D21S292
ACCESSION AP001599
VERSION AP001599
KEYWORDS AP001599.1 GI:7670553
SOURCE HTG.
Homo sapiens pre-pro-B cell cell_line: FLEB14-14 DNA, clone_11b: Keio
BAC library clone: KB126A3.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 118241)
AUTHORS Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE Homo sapiens genomic DNA, chromosome 21, clone: KB126A3, App-D21S292
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 118241)
AUTHORS Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
(E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370,
Fax: 81-3-3351-2370)
FEATURES
source
1..118241
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/chromosome="21"
/clone="KB126A3"
/clone_11b="Keio BAC library"
BASE COUNT 35861 a 22962 c 22601 g 36817 t
ORIGIN
Query Match 74.2%; Score 985.6; DB 9; Length 118241;
Best Local Similarity 88.5%; Pred. No. 8e-289;
Matches 1120; Conservative 0; Mismatches 122; Indels 23; Gaps 4;
Qy 1 tcaaaatcgaagagcttagaacttgctaacgcgcaaaagagggaacgtgttatctt 60
Db 63198 TCAAAATCGAAGAGCTTATGAAATGCTAATCATGACGAGAGGGGAACTTTTATTTT 63257
Qy 61 aggggaagaatgctgtatgttaatcaatctggaatcattactggaagaagttaaga 120
Db 63258 AGGGGAAGAATGCTTATTTATGTTAATCAATTCGGAATCATCAACCAAGAAATTAAGA 63317
Qy 121 aattgagatcgaatatagctagagcagaggaaccttaaaaa--cactgcaacctggggcc 179
Db 63318 AATTCAGATGAATACACGTAAGACAGAGAGCTTAAAAAACACTGAGACCTGGGGCC 63377
Qy 180 tctcagcgaatgagatgacctctccctctcttagagaccttaagcagctataatat 239
Db 63378 TCTCAGGCAATGATGATGCTCGATTCCTCCCTCTTATGAGACCTTACAGCT--ATAT 63434
Qy 240 tttaactctcttggagaccctgtaattctcaactctcctgttaagtgtgtctctccgaa 299
Db 63435 TTCCTACTCTCTTTGGACCCGTATCTTAACCTCCGCTTAAGTTGCTCTTCACAA 63494
Qy 300 ttgaagctgtaagctatacaatagttcttcaaatggaacccagatgagctcaatgacta 359
Db 63495 TCGAAGATGTAAACTACAAATCGTTCTTCAATGAGACCCCAAGATGACATGACTA 63554

```

QY 360 aaatctaccgtgagcccttgaccggcctgctagactatgctgtatgttaatatgct 419
 Db 63555 AGATCTACTGAGAGACCCCTTGACACGACCTCTGAGCCCATCTCCATGTTAATGACATTTG 63614
 QY 420 aagtcacccctcccgagaaatctcaactgacaaacccctactacactccattcagtag 479
 Db 63615 AAGGACACCCCTCCCAAGAAATCTCAACTGCACAAACCCCTACTATGCTCCAAATTCAGCAG 63674
 QY 480 gaagcaggttagagcagttgtcagccaaacctccccaaagtaactgtgtgttctgttgag 539
 Db 63675 GAAGCAGGTTTACAGTGTCTCCGGCCAAACCTCCCAACAGCATTTGATTTCTGTTGGG 63734
 QY 540 aggtgtgactgagagagacagactagctgtgattcccaagctgactaagaatccnaagc 599
 Db 63735 AGGGGGACCTGAGAGACAGAGACTAGCTGATTTCTTAAGCTGACTGAGAAATCCCTTAAGCC 63794
 QY 600 tancgtggaaggtgacgcacatcctttaacaatgagggttgcaacttagtcacaccc 659
 Db 63795 TACCTGGGAGAGTGACCACTTCCACCTTTAACAACAGGCGTTCGCAACTTGTAGCTCACACC 63854
 QY 660 gaccacatc-----agagagctcactaaatgctcactcaagcgaacaaacagaggttaa 710
 Db 63855 TACCAATTTGGATATGTAAGAGAGAGCTCAATAAATGCTAATTAGGCAAAAACAGAGGTAA 63914
 QY 711 agcaatagccaatcatctatgtcctgagagcagcagcggagagacaagatgtgatatata 770
 Db 63915 AGAATATAGCAATCATTCATTCGCTGAGACACAGCGGAGGAGCAATGACAGGATATA 63974
 QY 771 aactcaggaatcaagcagcagaacacacaccccttggtgctccctccatgtatagga 830
 Db 63975 AACCCAGGCAATTCACAGCTGACCAAGGCAACCCCTTTGGGTCCCTCTTCTTATAGGA 64034
 QY 831 gctcttcttcaactctatctatctatataatcagcaactgcaactctctgtgtcgtg 890
 Db 64035 GCTCTGT-----TTTACTCTATTCATATCTTGCACTGCACTCTTGTGCTCGTG 64084
 QY 891 tttttatggtcgaagctgagctgtgtgttcgcacatccacacactgtgtttgcacagctca 950
 Db 64085 TTTGTTACGGCTCAGCTGAGCTGTTTGTGCTCACCATCCACACAGCTGCTTGGCGCGCTTG 64144
 QY 951 cagaccgcgtgtgacttccatcccttttgatccagcagagatgtccactgtgtctcgat 1010
 Db 64145 CAGACCCGTGCTGACTTCCATCCCTCCAGATCTGGAGGGGTGCCACTGTGCTCCCTGAT 64204
 QY 1011 ccagcagagatcaccatggtccactcccgatcagcctaagccttgacatgttctcgatg 1070
 Db 64205 CCGCGAGGACCCCATTTGCGACTCCGATCAGGCTAAAGGCTTGCCATTTCTCTGCACA 64264
 QY 1071 gctaaagtgcctgggtgtgttctcctaatagaactggaacactgtgctggttccatgttct 1130
 Db 64265 GCTAAGTGCTGGGTTGCTGCTAATCAAGCTGAACACTAGTCACTGGGTTCCATGGTCTC 64324
 QY 1131 ctccatgacccacagcgtcttataatagacataaacactcacgcgatggcccaagatccca 1190
 Db 64325 CTTCACATGACCCATGCTTTAATAGAGCTAATAACATCAACCGCATGGCCCAAAATTTCCA 64384
 QY 1191 ttcctgtgtctcgtgagcgaagaacccaggtcagagaangtgaagcttgccacacatt 1250
 Db 64385 TTCTCTTGGAATCCGTGAGGCGCAAGACCCCAAGGTGAGAAACAGAGGCTGCGCCCATCT 64444
 QY 1251 tggga 1255
 Db 64445 TGGAA 64449

RESULT 12

AP001697

LOCUS AP001697 340000 bp DNA linear PRI 30-MAY-2000

DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 41/105.

ACCESSION AP001697 AL163242 BA000005

VERSION AP001697.1 GI:7768705

KEYWORDS

SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
 1 (sites)
 Hattori,M., Fujiyama,A., Ishii,K., Tokok,Y., Choi,D.K., Soeda,E.,
 Park,H.S., Toyoda,A., Ishii,K., Taudien,S., Blechschmidt,K.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., and Schudy,A.
 The DNA sequence of human chromosome 21. The chromosome 21 mapping
 and sequencing consortium
 Nature 405 (6784), 311-319 (2000)
 20289799

JOURNAL Nature 405 (6784), 311-319 (2000)
 MEDLINE 20289799
 REFERENCE 2 (bases 1 to 340000)
 AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
 Park,H.S., Toyoda,A., Ishii,K., Tokok,Y., Choi,D.K., Soeda,E.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
 Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
 Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstieck,G.,
 Hornischler,K., Barand,P., Scharfe,H., Schoen,O., Desario,A.,
 Reichel,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
 Hennig,S., Rieseemann,L., Dagnand,E., Wehmeier,S., Borzym,K.,
 Gardner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
 Vaspo,M.L.

TITLE Direct Submission
 JOURNAL Submitted (10-Apr-2000) The Chromosome 21 Mapping and Sequencing
 Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
 Group * Institute of Molecular Biotechnology, Genome Analysis *
 Keio University School of Medicine, Dept. of Molecular Biology *
 GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
 Genetics (addresses see below)

COMMENT

On May 30, 2000 this sequence version replaced gi:7717295.
 The chromosome 21 mapping and sequencing consortium consisting of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagami-hara 228-8555, Japan,
 * e.mail: hattori@gscl.riken.go.jp/
 * URL: http://hnp.gsc.riken.go.jp/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *
 Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Molecular Biology, * Tokyo
 160-8582, Japan,
 * e.mail: nshimizudmb-med.keio.ac.jp/
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
 info.genome@gbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Inneistrasse 73, D-14195 Berlin, Germany,
 * e.mail: info-chr21emo@gen.mpg.de
 * URL: http://chr21.fz-berlin.mpg.de/
 AL163242: Submitted (10-Apr-2000).
 Location/Qualifiers
 1..340000
 1..340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 .1..75790
 /note="Accession No. AP001597"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q21.2"

FEATURES

source


```
/clone="KB1951E10, 5' partial"
/clone_id="Keio BAC library"
737. .1092
/repeat_region
/note="HERV.L"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
complement(1763. .2230)
/note="MLT1D"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(3756. .4210)
/note="Charlie4a"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
4322. .4476
/note="L1P45"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
5482. .5758
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
5759. .5866
/note="(GCAN)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
8386. .8557
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
8822. .9116
/note="AluY"
/rpt_family="SINE/AluY"
/rpt_type=DISPERSED
10343. .10460
/note="(TTCG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(10463. .10632)
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(10635. .10669)
/note="Alu"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(10852. .11010)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
11094. .11142
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(12123. .12810)
/note="LINE"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
12952. .12985
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(13168. .13419)
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(13461. .13746)
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
14398. .14604
/note="L1M2"
/rpt_family="LINE/L1"
```

```
/rpt_type=DISPERSED
14604. .14753
/note="L1M4C"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
14810. .15161
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(15167. .16789)
/note="MER52A"
/rpt_family="LTR/MER4_group"
/rpt_type=DISPERSED
16799. .17007
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(17139. .17448)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
19049. .19312
/strand.name="D2LS269"
/note="AFM263xf5"
Accession No. Z17183"
19105. .19159
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(19358. .20123)
/note="L1M49"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(20142. .20413)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(20415. .20792)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
20863. .20889
/note="(TC)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(20970. .21075)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED

Query Match 74.2% Score 985.6; DB 9; Length 340000;
Best local similarity 88.5%; Pred. No. 8.1e-289;
Matches 1120; Conservative 0; Mismatches 122; Indels 23; Gaps 4;

QY 1 tcaaatcgaaagccttagacttgctaacgcgcaaaagagggaacctgttatctt 60
|||||
Db 250910 TCACAATCGAAGACCTTAGAATGCTAATACGAGAGGAGGAGCGTTTATTTT 250969

QY 61 aggggaagaatgcgttagttagtcaatcaatcggaaatcaatcctgagaaagttaaga 120
|||||
Db 250970 AGGGGAAGAATGCGTTRTATTTATGTTATCAATTGCGAATCATCCACGAAGAAATTAAAGA 251029

QY 121 aattgagatcgaaataatgtagagcagagagacctcaaaa-cactgcacctggggcc 179
|||||
Db 251030 AATTCAGATCGAATFACAACTGACAGAGAGCTTAAACACAGACCTGGGGCC 251089

QY 180 tcccaagcgaatgtagtccctggagctctccctccttagagaccctgagcgtataat 239
|||||
Db 251090 TCCACAGCAATGATGCGCCGGAATCTCCCTCTTAGAGACCTAGCAGACT--ATAT 251146

QY 240 tttaactcctcttgagaccctgtagtcaactcctgtagttagtctctccaga 299
|||||
Db 251147 TTCTACTCCTCTTGGACCCCTGTATCTTTAACTCCGCTGTTAGTTGTCTTCCAGAA 251206
```


RESULT	13		
AP001600			
LOCUS	91097 bp	DNA	linear
AP001600			PRI 29-APR-2000

DEFINITION	Homo sapiens genomic DNA, chromosome 21, clone:KB45E1, App-D2JS292
ACCESSION	region, complete sequence. Ap001600
VERSION	Ap001600.1 GI:7670554
KEYWORDS	HTG.
SOURCE	Homo sapiens pre-pro-B cell cell_line:FLB14-14 DNA, clone_lib:keio BAC library clone:KB45E1.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE	1 (bases 1 to 91097)
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE	Homo sapiens genomic DNA, chromosome 21, clone:KB45E1, App-D2JS292
JOURNAL	region
REFERENCE	Published Only in Database (2000) In press
AUTHORS	2 (bases 1 to 91097)
TITLE	Shimizu,N., Kudoh,J. and Shibuya,K.
JOURNAL	Direct Submission
REFERENCE	Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.
AUTHORS	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan (E-mail:shimizu@med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES	
source	Location/Qualifiers 1..91097 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="FLB14-14" /cell_type="pre-pro-B cell" /chromosome="21" /clone="KB45E1" /clone_lib="Keio BAC library"
BASE COUNT	28309 a 16954 c 17419 g 28415 t
ORIGIN	
Query Match	74.0%; Score 984; DB 9; Length 91097;
Best Local Similarity	88.5%; Pred. No. 2.4e-288;
Matches 1119; Conservative	0; Mismatches 123; Indels 23; Gaps 4;
QY 1	tcaaatcgaaagcgttagacttgcctacgcgcgaagaggagggaacctgttatctttc 60
DB 9838	TCAAATCGAAGAGCTTTAGAAATTCCTATATCTGATGAGAGGGGAGACGTTTATTTT 9917
QY 61	aggggaaagatcgcttagtatgttaactcaactcgtgaatactatctagaagaagtaaga 120
DB 9918	AGGGAGAAGATCTCTTATTATGTTTAATCAATTGCGAATCATCACCAAGAAAGTTAAAGA 9977
QY 121	aatttgatcgaatataatgtatgagacgagagaccctcaaaa-cactgcaccttgaggcc 179
DB 9978	AATTCAAGATCGAATATACAACGTAGAACAGAGAGAGCTTAAAAAACACTGAGACCTGGGGCC 10037
QY 180	tecttaagccaatggatgcccctggaactcccccctctttgagaccttagcagctataatc 239
DB 10038	TCCTTAGCCAAAGATGCGCCCTGAGATTCTCCCTCTTAGGAGACTTACAGACT--AATAT 10094
QY 240	tttaactcctcttggaacctgtatcttcaacttaactctctgttaagtttgccttcacaga 299
DB 10095	TTTCTACTCTCTTGTGGACCTGTATCTTATTAACCTCGGTTAAGTTGTCTTCCAGAA 10154
QY 300	tttgaagctgtaagctacaatagttcttcaatgtaagaaacccagatgcagtcacatgacta 359
DB 10155	TCGAAGATGTAAACTACAAATCGTTCTTCAAAATGAGACCCCGACAGTGCATGATGACTA 10214
QY 360	aaatctaccgtygagccctgagccgagcctgctagactatgctctgattgtaatgacatg 419
DB 10215	ACATCTACGTAGAGAGCCCGTGAGACGCGGCGTACGCCCATGCTCAATGTAAATGACATTG 10274
QY 420	aagtaacccctccgaggaagaatctcaactgtgacaaacccctatacactccaattcaagtag 479
DB 10275	AAGGACACCCCTCCCAAGGAAGAACTCAACTGCAACACCCCTACTATGCTTCCAAATTTAGAGC 10334
QY 480	gaagcaqitlaagcaqitltcagccaacccctcccaacatctatgttgatttctctgttag 539

```

Db 10335 GAACGAGTTACAGTGTCTCTGGCCAACTCCCAACAGCATTTGTATTTCTCTGGG 10394
Qy 540 agggctgacatgagacagacagactagctgattctcttaagctgactaaatcccaagcc 599
Db 10395 AGGGGCGACTGAGACAGAGACTAGCTGATTTCTCTAGGCTGACTGACAAATCCCTAAACC 10454
Qy 600 tancctggaagtgagccgcatcctcttaacacagtgagcttgaaacttagctacaccc 659
Db 10455 TAGCTGGGAGAGTGACCACTTCACTTTAAACAGAGGCTTGCAACTTACTGACACACC 10514
Qy 660 gaccacatc-----agaaagctcactaaatgctctaaagcgaataaagaagagtaa 710
Db 10515 TACCAATTGGATAGTAAAGAGAGGCTAATAAATGCTAATTAGCCAAAACAGAGAGTAA 10574
Qy 711 agacatagacatcatctatctgctcgaagacagcagcgaagaaagacaaagtgatata 770
Db 10575 AGAATATACCAATCATCTTCATTTGCTGAGAGCAGACCGGAGGACAAATGACAGAGATTA 10634
Qy 771 aactcagagcatcacaagcacaagacacccctcttggtccctccctccatctgatagga 830
Db 10635 AACCCAGGCAATTCAGACCTGCAAGCGCAACCCCTTTGGTCCCTCTCTTGTATGGGA 10694
Qy 831 gctctgtttcactcctattcactcctataaatacagctgactctcttctgtcgtg 890
Db 10695 GCTCTGT-----TTTCACTCTATTCAATCTTTCACACTGCACTCTTGTGTCGTG 10744
Qy 891 tttcttaagctcacaagctgagctcttgctgcacatccacactgctgcttgacagctca 950
Db 10745 TTGTGTACGCGCTCAAGCTGAGCTTTTGCTGACAACTGCACTGCTGTGTTGCCCGCTG 10804
Qy 951 cagacccgctgctgactcactcctcttgatcgaagagtgctccactgctgctcgtat 1010
Db 10805 CAGACCCGCTGCTCACTCTCTCTCCCTCCAGATCGGCGGCTGCTGCTGCTGCTGAT 10864
Qy 1011 ccagcgaagtgacatgacatgacatccctcgaatcagctgactgacatgctgctcgtat 1070
Db 10865 CCAGCGAGGCAACCATTCCTCCCTCCATTCAGGCTTAAAGGCTTGCTGCTGCTGCA 10924
Qy 1071 gctaaagctgctgctgctgctcctaaatagaactgaactgctgctgctgctgctct 1130
Db 10925 GCTAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10984
Qy 1131 ctccatgagacacagctctcctaaatagaactgaactgaactgctgctgctgctgct 1190
Db 10985 CTTTCATGACCCATGCTCTTAATAGACCTTAACACCTGACCCGATGGCCCAATTTCCA 11044
Qy 1191 ttcctgtgctatctgctgagcgaagaaacccagctcagaagaangtgagcgtgctgacat 1250
Db 11045 TTCCCTTGGATCCGTGAGGCCAAGACCCCAAGTCAAGAACAGAGGCTGCCCATCT 11104
Qy 1251 tggga 1255
Db 11105 TGGAA 11109

```

```

RESULT 14
AC008121 172721 bp DNA linear HTG 16-OCT-2001
LOCUS Homo sapiens chromosome 12 clone RP11-407N8, WORKING DRAFT
DEFINITION AC008121
SEQUENCE 20 unordered pieces.
AC008121.25 GI:14190616
VERSION AC008121.25
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 172721)
AUTHORS Muzny,D.M., Adams,C., Adio-Ogunola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

```

TITLE JOURNAL AUTHORS REFERENCE JOURNAL

COMMENT

```

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dabholkar, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C.,
Elhaj, J., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gilly, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,
Hollins, B., Homsi, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Licheare, O., Lieu, C., Liu, J., Liu, W.,
Louiseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabpat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Oguni, M., Okunola, G., Oragunye, N., Oriedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, K., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Stinson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Swalek, A., Taber, P., Tamejia, A., Tamejia, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczek, R., Woodson, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 172721)
Worley, K.C.
Direct Submission
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 23, 2001 this sequence version replaced gi:13940575.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMR
Center clone name: RP11-407N8
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 41% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 152097 bases at least Q40
Consensus quality: 166730 bases at least Q30
Consensus quality: 173153 bases at least Q20
Estimated insert size: 173988; sum-of-ctrls estimation
Quality coverage: 0x in Q20 bases; agrose-fp estimation
Quality coverage: 3x in Q20 bases; sum-of-ctrls estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```


RESULT 15
AC016173/c
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-440C21 map 2. WORKING DRAFT
AC016173 145785 bp DNA linear HTG 27-APR-2000
SEQUENCE, 17 unordered pieces.
AC016173
VERSION AC016173.2 GI:7657740
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE (bases 1 to 145785)
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 145785)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeRubeis, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, D., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lemock, J., Lieu, C., Locke, R., MacDonald, P., Marquis, N.,
McManis, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teafaye, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wymann, D., Ye, W., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 27, 2000 this sequence version replaced gi:6466621.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L3813
Center clone name: 440_C-21
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136152 bases at least Q40
Consensus quality: 140073 bases at least Q30
Consensus quality: 141708 bases at least Q20
Insert size: 18700; agarose-fp
Insert size: 14185; sum-of-ctnigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctnigs
NOTE: This is a 'working draft' sequence. It currently
consists of 17 ctnigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the ctnigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1331: contig of 1331 bp in length
* 1332 1431: gap of 100 bp
* 1432 2628: contig of 1197 bp in length
* 2629 2728: gap of 100 bp
* 2729 3745: contig of 1017 bp in length
* 3746 3845: gap of 100 bp

FEATURES

source

3846 6327: contig of 2482 bp in length
* 6328 6427: gap of 100 bp
* 6428 7950: contig of 1523 bp in length
* 7951 8050: gap of 100 bp
* 8051 10578: contig of 2528 bp in length
* 10579 10678: gap of 100 bp
* 10679 15308: contig of 4630 bp in length
* 15309 15408: gap of 100 bp
* 15409 20910: contig of 5502 bp in length
* 20911 21010: gap of 100 bp
* 21011 25032: contig of 4022 bp in length
* 25033 25132: gap of 100 bp
* 25133 29129: contig of 3997 bp in length
* 29130 29229: gap of 100 bp
* 29230 36550: contig of 7321 bp in length
* 36551 36650: gap of 100 bp
* 36651 48963: contig of 12313 bp in length
* 48964 49063: gap of 100 bp
* 49064 62673: contig of 13610 bp in length
* 62674 62773: gap of 100 bp
* 62774 75359: contig of 12586 bp in length
* 75360 75459: gap of 100 bp
* 75460 90704: contig of 15245 bp in length
* 90705 90804: gap of 100 bp
* 90805 109887: contig of 19083 bp in length
* 109888 109987: gap of 100 bp
* 109988 145785: contig of 35798 bp in length.
Location/Qualifiers
1. 145785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-440C21"
/clone_lib="RPC1-11 Human Male BAC"
1. 1331
/note="assembly_fragment"
1432. 2628
/note="assembly_fragment"
2729. 3745
/note="assembly_fragment"
3846. 6327
/note="assembly_fragment"
clone_end:17
vector_side:right
6428. 7950
/note="assembly_fragment"
8051. 10578
/note="assembly_fragment"
10679. 15308
/note="assembly_fragment"
15409. 20910
/note="assembly_fragment"
21011. 25032
/note="assembly_fragment"
25133. 29129
/note="assembly_fragment"
29230. 36550
/note="assembly_fragment"
36651. 48963
/note="assembly_fragment"
49064. 62673
/note="assembly_fragment"
62774. 75359
/note="assembly_fragment"
clone_end:SP6
vector_side:left
75460. 90704
/note="assembly_fragment"
90805. 109887
/note="assembly_fragment"
109988. 145785
/note="assembly_fragment"

BASE COUNT 4438 a 27590 c 27266 g 44991 t 1600 others
 ORIGIN

Query Match 73.7%; Score 979.2; DB 2; Length 145785;
 Best Local Similarity 88.2%; Pred. No. 7.2e-287;
 Matches 1116; Conservative 0; Mismatches 126; Indels 23; Gaps 4;

QY 1 tcaaatcgaagagcttgaacttgcctgaaccgcaaaagggggggaacccgtttatttt 60
 DB 7898 TCAAAATCGAAGACTTTAAATTTGCTAAATCACTGAGAGAGGGGAACTTTTATTTT 7839
 QY 61 aggggaagaatgctgttagatgttaatcaatcctggaatcaatctactgaagaagttaaaga 120
 DB 7838 AGGGGAAGATGCTGTATATATGTATATCAATTCGGAATCATCAACCAAGAAAGTTAAGA 7779
 QY 121 aatttgatcgaataataatgttagagcagagacccctcaaaa-cactgcaccttgaggcc 179
 DB 7778 AATTCAAGATCGAATACAGTGAACAGAGAGGAGCTTAAAAAACACTGAGACCTGGGGCC 7719
 QY 180 tccctcaagcaatgagtcgcccctgagacccctccctctcttagagcccttagcagctaaat 239
 DB 7718 TCCTCAGCCAAATGAGATGCTGGATTTCTCCCTTTTGAAGCTCTAGCAAGCT--ATAT 7662
 QY 240 tttaactcctcttggaccctgtatcttcaactccttctgttaagtcttctctccagaa 299
 DB 7661 TTCTACTCCTTTTGGACCTGTATCTTTAACCTCGCTGTAAAGTTGTCTCTCCGAA 7602
 QY 300 ttgaagctgtaagctacaataagttcttcaaatgaaacccagatgacgtacata 359
 DB 7601 TCGAAGATGTAAACTCAATATGCTTCTCAAAATGAGACCCCAAGATGACGTGACTA 7542
 QY 360 aaatcaccgtgagacccctgagacccctgagacccctgagacccctgagacccctgagac 419
 DB 7541 AGATCTACTAGAGACCCCTGAGACCCCTGAGACCCCTGAGACCCCTGAGACCCCTGAG 7482
 QY 420 aagtcacccctcccgagaaatctcaactgacacacccctactacactcaatcaatag 479
 DB 7481 AAGGCACCCCTCCCAAGGAATCTCAACTGACACACCCCTACTATGCTCCAAATTCAGCAG 7422
 QY 480 gaagcagttgagcagttgttcaagcaaacctcccaacagtaacttgggttttctcgttag 539
 DB 7421 GAAGCAGTTGACAGTGTCTGAGCCCAACCTCCCAACAGATTGTATTTTCTGTTGGG 7362
 QY 540 aggtgagcag 599
 DB 7361 AGGGGACACTGAGAGACAGAGACTGAGATTTCTTAGGCTGAGACTGAGAAATCCCTAAAGCC 7302
 QY 600 tancctggaaagtgagccgcaactcttcaacaatggggttgcacttgaacttgaactcaca 659
 DB 7301 TAGCTGGGAAGGTGACCACTTCACTTTTAAACAGAGGGCTTGCACCTTAGCTCACACC 7242
 QY 660 gacccaatc-----agagagctcaataatgtcaatcagagcaaaaacagaggtaa 710
 DB 7241 TACCAATTGATAGTAAGAGAGAGCTCAATAATGCTAATTAAGCAAAACAGAGAGTAA 7182
 QY 711 agcaatagcaatcatatcttctgagagcagcagcagcagcagcagcagcagcagcagcag 770
 DB 7181 AGAAATAGCCAAATCATCTTCTGAGAGCAGCAGCGGAGGAGCAATGACCAAGATATA 7122
 QY 771 aactcagagcattcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 830
 DB 7121 AACCAAGGCAATTCAGGCTGAAAGGCAACCCCTTTGGGTCCCTCTTTTATGGA 7062
 QY 831 gctcgttttcaactctatctcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 890
 DB 7061 GCTCTGT-----TTTCACTCTATCTCAATCTTGCACACTTCTGCTGCTG 7012
 QY 891 ttttttggctcaagcttggcttcttggcactcagcagcagcagcagcagcagcagcagcag 950
 DB 7011 TTTTGTAGGGCTCAAGCTGAGCTTTTGGCTCAGCACTCAGCTCTGTTTGGCCGCTTG 6952
 QY 951 cagaccgctgtgacttccacttggatccagcagagtgctcaactgtctccttgaat 1010

DB 6951 CAGACCCGCTGCTGACTTCCATCCCTCCAGATGCGGAGGGTGTCCACTGTGCTGAT 6892
 QY 1011 ccagcagagtaaccatctgcaactcccgatcagcagcagcagcagcagcagcagcagcag 1070
 DB 6891 CCAGCGAGGACCCATGCTGCGCTCCCAATGCGGCTAAAGCCTTGCATTTGCTTGCACA 6832
 QY 1071 gctaagtcgctgggttgccttaataagaaactggaactggttcaactggttccatggtctc 1130
 DB 6831 GCTAAGTCCCTGGGTTGCTCTAATCAAGCTGAACACTAGTCACTGGGTTCCATGTTCT 6772
 QY 1131 ctccatgacccacgcttcttaataagcctataaacactacccgaaatgcccagaatcca 1190
 DB 6771 CTTCATGACCCATGCTTCTTAATAGAGCTATTAACACTCAACCGCATGGCCCAAGATTCA 6712
 QY 1191 ttccttggatctgttggagccagaacccagcagcagcagcagcagcagcagcagcagcag 1250
 DB 6711 TTTCTTGGAAATCCGTGAGGCCAAGAACCCAGGTCAAGAAACAGAGGCTGTGCTCATCT 6652
 QY 1251 tggga 1255
 DB 6651 TGGAA 6647

Search completed: June 20, 2002, 10:15:40
 Job time: 12665 sec

THIS PAGE BLANK (USPTO)